STRUCTURAL AND THERMODYNAMIC CONSEQUENCES OF INTERNAL POLAR AND IONIZABLE RESIDUES IN STAPHYLOCOCCAL NUCLEASE

by
Jaime Lynn Sorenson

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
April 2016

Abstract

Ionizable residues govern many biological processes including energy transduction and enzymatic reactions. Ionizable groups are highly sensitive to their environment and have different properties when in the presence of or sequestered from bulk solvent. Understanding the factors that determine pK_a values of the internal ionizable residues will greatly improve our ability to predict the behavior of proteins with natural or engineered ionizable groups.

Previous work with staphylococcal nuclease (SNase) with substitutions to Asp, Glu, Lys, and Arg residues at 25 different internal positions showed that the pK_a values can be shifted up to 5.5 pK_a units relative to the model compound value. This thesis investigates the determinants of pK_a values and the effects that polar and ionizable residues have on protein structure and stability. By placing a single His residue at 25 different internal locations in SNase, we describe the influence of the microenvironment on the pK_a and illustrate the unique behavior of His. Further studies with Lys at similar positions show the extent to which global protein stability affects pK_a values. The results of these experiments will have a great impact on structure-based computational algorithms used to predict pK_a values in proteins. Lysine was also used to validate a novel approach towards the systematic engineering of a protein to have high sensitivity to pH with the goal of undergoing a global unfolding event at a specified pH threshold. The findings presented in this work will provide insight beneficial to the fields of bioenergetics and biopharmaceuticals.

Thesis Committee

Bertrand García-Moreno E., Ph.D. (Advisor, Reader)

Vince Hilser , Ph.D. (Reader)

Juliette Lecomte, Ph.D.

Gregory Bowman, Ph.D.

Acknowledgments

I would like to thank my advisor Dr. Bertrand García-Moreno E. for his support and guidance and for accepting me into his lab even though I had no prior biophysics experience. I am grateful to past members of the BGME lab for laying the scientific foundation for my work, particularly Dr. Carlos Castañeda (for chats about surviving grad school and swing dancing) and Erika Wheeler (for being the other half of Jaimerika). Thanks to the current members for their continuing advice and creating an entertaining and friendly work environment, especially Dr. Aaron Robinson (for always knowing something about everything), Peregrine Bell-Upp (for being a great presence in the lab and at the boathouse), Dr. Dan Richman (for scientific and pedagogy discussions), and Dr. Jamie Schlessman (for teaching me everything about crystallography). Thanks also to the Lecomte lab for making the basement bright even with the lack of windows: Dr. Matt Pond, Dr. Matt Preimsberger, Dr. Eric Johnson, Dr. Selena Rice, and Belinda Wenke.

I also wish to thank the members of the Biology and Biophysics departments for thoughtful questions about my projects. I am grateful to the past and present members of MInDS who became some of my first and closest friends in the Biology department. Thanks to Diamond Ling and Chase Hermsen for being close friends and roommates since Rochester and Katie Svilar Griebel and Elizabeth Larner for the support through tough times and the exchange of snail-mail letters.

I would like to thank my family, both human and animal, for the immense support. Thanks to my mom (Michelle) for always being there, my brother and his wife (Josh and Yvonne) for trying to understand the more abstract parts of my projects, and my sister (Jacquie) for the great fashion tips and always willing to talk.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Most polar and ionizable amino acids are found on the protein-water interface, where they can interact with water. A few of them are buried in the hydrophobic and dehydrated interior of the protein. Owing to their dehydration, these buried polar and ionizable groups can have highly anomalous properties and they can affect the stability and dynamics of a protein significantly. The few buried polar or ionizable groups that are found in proteins are usually essential for biological processes. One of the goals of this dissertation is to examine the relationship between the presence of polar and ionizable residues in the hydrophobic interior and the stability and conformational states of proteins. It is important to understand how these groups contribute to protein function, to understand the potentially deleterious consequences of spontaneous mutations of internal hydrophobic residues into polar or ionizable residues, and to learn how to harness the atypical properties of buried polar or ionizable groups for the engineering of novel protein function. Another goal was to contribute to the deeper understanding of the physical properties of proteins that determine their behavior as a dielectric material when exposed to charged groups. This is important because this behavior governs the magnitude and character of electrostatic effects involving these buried groups. We anticipate that the results of the experiments discussed within this dissertation will also be useful to guide improvements for computational methods for structure-based energy calculations and as benchmarks to examine the validity of these calculations.

Results from four different studies are described in this dissertation: (1) Thermodynamic stability measurements to determine consequences of substitution of buried residues in staphylococcal nuclease (SNase) with polar residues with hydroxyl groups (Ser, Thr, Tyr). To our knowledge, this constitutes the first systematic study of

hydrophobic to polar substitutions at buried positions in proteins. (2) Determination of the effects of pH on the thermodynamic stability of variants of SNase with His at internal positions. This study complements ongoing research in this laboratory on the anomalous properties of buried Asp, Glu, Lys, and Arg residues. The goal was to determine if His residues behave like Arg residues, which are charged even when buried in internal position[1], or like Lys residues, which are usually neutral when buried[2]. (3) Measurement of the effects of global thermodynamic stability on the pK_a values of an internal Lys residue with a highly anomalous pK_a value. The goal was to examine the hypothesis that the probability of conformational rearrangement concomitant with ionization of a buried group is the most important determinant of the atypical pK_a values of buried ionizable residues. (4) Demonstration that buried Lys residues with anomalous pK_a values are useful to engineer proteins that act as pH sensitive switches responsive to small changes in pH in the physiological range.

1.1 Polar and ionizable residues buried in the hydrophobic interior of a protein

1.1.1 Serine, threonine, and tyrosine

Three categories of polar amino acids can be distinguished based on the identity of the side chain: asparagine (Asn) and glutamine (Gln) have amides, cysteine (Cys) and methionine (Met) have sulfur, and serine (Ser), threonine (Thr), and tyrosine (Tyr) have a hydroxyl group. Asn, Gln, Ser, Thr, and Tyr are very hydrophobic. In this dissertation, we focus exclusively on the properties of internal amino acids with hydroxyl groups on the side chains: Ser, Thr, and Tyr.

Polar residues are normally found at the surface of the protein where they can interact with water[3]. In general, the burial of a polar group in the hydrophobic interior of a protein without inclusion of a hydrogen bonding partner is destabilizing[4]. Polar side chains can mitigate the potential destabilizing consequences of being buried by forming hydrogen bonds to the protein backbone, other polar side chains, or internal water molecules. The magnitude of the destabilization of the native state related to the burial of polar groups can be proportional to the polarity of the amino acid. It appears that the dehydration cost can be counteracted if the polar residue forms hydrogen bonds to other nearby polar groups[5,6], but this has never been examined systematically; this is one of the goals of this dissertation.

Buried polar residues are known to take on a variety of roles in biological processes including catalysis[7,8] and cellular signal propagation[9,10]. In enzymes, 27% of the catalytic residues involve polar residues[7]. For example, a hydroxyl group can form hydrogen bonds and act as a nucleophile in chemical reactions[7,8,11] or act as a site of covalent modification through phosphorylation[12,13]. Besides direct action on a substrate as a nucleophile, the hydroxyl group can act indirectly by activating catalytic water molecules, cofactors, or stabilizing the transition state[7].

1.1.2 Ionizable residues: Asp, Glu, Lys, Arg, and His

Many structural, physical, and functional properties of proteins are governed by acidic and basic ionizable residues. Buried ionizable residues are necessary for many biological processes including enzyme catalysis[14–17], H^+ transport and e^- transfer[18–21], ligand coordination[22–27], regulation of allostery and kinetics[28–30], and ion homeostasis[31,32].

In water, the side chains of aspartic acid (Asp) and glutamic acid (Glu) have pK_a values of approximately 3.9 and 4.4, respectively[33,34]. When a carboxylic side chain is buried in the hydrophobic interior of a protein, their pK_a values are usually shifted towards higher values, favoring the neutral state[35,36]. This shift in pK_a is reflected in the pH-dependence of the thermodynamic stability of a protein ($\Delta G°_{H2O}$).

When ionized, surface Asp residues tend to disrupt local or global structure more than Glu, consistent with the stronger preference of Asp for unstructured regions of proteins[37,38]. In a survey of Asp and Glu at 25 internal positions in SNase, 4 out of 25 Asp-containing variants were globally unfolded when the Asp residue was charged and only 1 of the 25 Glu-containing variants exhibited the same behavior. Similarly, 11 of the variants with Asp mutations showed local structural changes while only 4 of the variants with Glu showed the same. Presumably, the local structural changes triggered by the ionization of these buried residues somehow increased the solvation of the ionizable group. Such reorganization coupled with ionization of internal ionizable residues is difficult to reproduce with structure-based electrostatics calculations[39].

The side chains of Lys, Arg, and His, in water, have pK_a values of approximately 10.4, 12, and 6.5, respectively[33]. When buried a hydrophobic region of a protein, Lys can have drastically depressed pK_a value[2]. The pK_a values of Lys residues at 25 internal positions in SNase ranged from 10.4 to 5.3[2]. In contrast, a study of Arg at the same 25 positions failed to detect any shifts in pK_a value. This suggests that Arg always find a way to completely satisfy its hydration and hydrogen bonding potential, even when buried at internal position[1]. More recently it was shown that the pK_a value of the guanadinium moiety of Arg in water is probably greater than 14, much higher than the

value of 12 that is quoted in many textbooks[40].

The side chain of histidine has a pK_a value near 6.5 in water[33,41]. The imidazole side chain is complex as it is asymmetric and has some aromatic character. In the neutral state, histidine has two tautomeric states. The two imidazole nitrogen atoms have different affinities for H^+ and therefore have different pK_a values[42,43]. Any accurate structure-based calculations of the pK_a of a His residue must calculate the correct relative populations of the tautomers. These calculations are further complicated because His can act as both a hydrogen bond donor or acceptor[25,44]. An accurate model of the His side chain is necessary to account for the role of H-bonding in determining its pK_a. The aromatic nature of the imidazole group allows it to interact with other aromatic or charged residues in parallel or orthogonal orientations through π-electron interactions[25]. Although structure-based calculations of the pK_a values of Lys and Arg are far from simple, calculations with the His side chain remain a formidable challenge.

1.2 Previous studies of internal ionizable groups in SNase

Most proteins that use internal ionizable residues for biological functions are large, complex, and unwieldy, making them too difficult to study with experimental biophysical methods. SNase, a small globular protein from the bacterium *Staphylococcus aureus*, is a well-characterized protein[45–48] that has been used to study the molecular determinants of pK_a values for many internal groups. SNase is a monomeric protein that is particularly well-behaved in equilibrium thermodynamic experiments, highly soluble and useful for NMR spectroscopy experiments, and relatively easy to crystallize and amenable to x-ray diffraction studies. It has a defined hydrophobic core which is ideal for studying the

effects of burial of polar and ionizable residues. In addition, site-directed mutagenesis has been used to increase the thermodynamic stability of the protein from 5.4 to nearly 14 kcal/mol, and the stability is relatively invariant from pH 4 to pH 10[49]. This is important for our studies because the goal is to measure the effects of polar and ionizable groups on thermodynamic stability and this would be impossible to achieve if the internal groups were not tolerated by the protein. SNase is an ideal protein in which to examine the molecular determinants of pK_a values of internal ionizable residues and the cost of substitution of internal positions with polar and ionizable groups.

1.2.1 Surface His, Asp, and Glu residues

The properties of ionizable residues in wild type SNase have been characterized extensively by NMR spectroscopy. SNase contains four His residues. His-8 and His-46 are fully exposed on the surface of the protein with pK_a values measuring 6.5 and 5.9 respectively[50]. His-121 and His-124 are involved in extensive hydrogen bonding networks and are subject to unfavorable Coulomb interactions by nearby ionizable groups[50]. The pK_a values of these two histidines measured by NMR spectroscopy are depressed to 5.3 and 5.7, respectively. A similar set of experiments using multidimensional and multinuclear NMR was undertaken to examine the pK_a values of surface Asp and Glu side chains. The pK_a values of 13 of the 20 surface Asp and Glu side chains were depressed by at least 0.5 units and only 6 were relatively unchanged from the model pK_a values[34]. There was only one residue, Asp-21, that was significantly elevated to a pK_a of 6.5, which is highly unusual and is a result of the intricate network of interactions between residues in the active site of the protein[34,51].

1.2.2 Internal ionizable residues: Asp, Glu, Lys, and Arg

This laboratory previously measured pK_a values for Asp, Glu, Lys, and Arg buried, one at a time, at 25 internal positions in SNase[1,2,35,36]. None of the 25 Arg-containing variants show an appreciable shift in the pK_a value (ΔpK_a) of the Arg, but 61 of the 75 other variants with Lys, Glu, or Asp substitutions had measurable ΔpK_a values, all in the direction that favored the neutral state. Compared to the normal pK_a values of Asp, Glu, and Lys in water, the magnitude of ΔpK_a ranged from 0 to 5.7 pK_a units (Fig. 1.1). Ninety three of the 100 variants tolerated the presence of an internal ionizable residue without major structural rearrangement determined by CD spectroscopy, however 7 variants (Glu-92, Lys-92, Lys-100, Asp-92 Asp-99, Asp-103, Asp-104) showed global or nearly global unfolding coupled to the ionization of the internal group[2,35,36,52].

The goal of those studies was to characterize the magnitude of ΔpK_a and the impact of the buried ionizable group on the thermodynamic stability of the protein. The instances in which the ionization energy of the residue was greater than the global stability of the protein and in which the protein was unfolded leading to a breakdown of the protein as a dielectric material were not useful towards understanding how the microenvironment of an ionizable group affects its pK_a. On the other hand, those variants provided clear evidence that conformational transitions that expose the previously buried ionizable group to water could affect the pK_a value[52]. How the structure of the protein changes as a result of the ionization of the buried group was the focus of a subsequent NMR spectroscopy study[53] and is an underlying motivation behind multiple chapters of this dissertation.

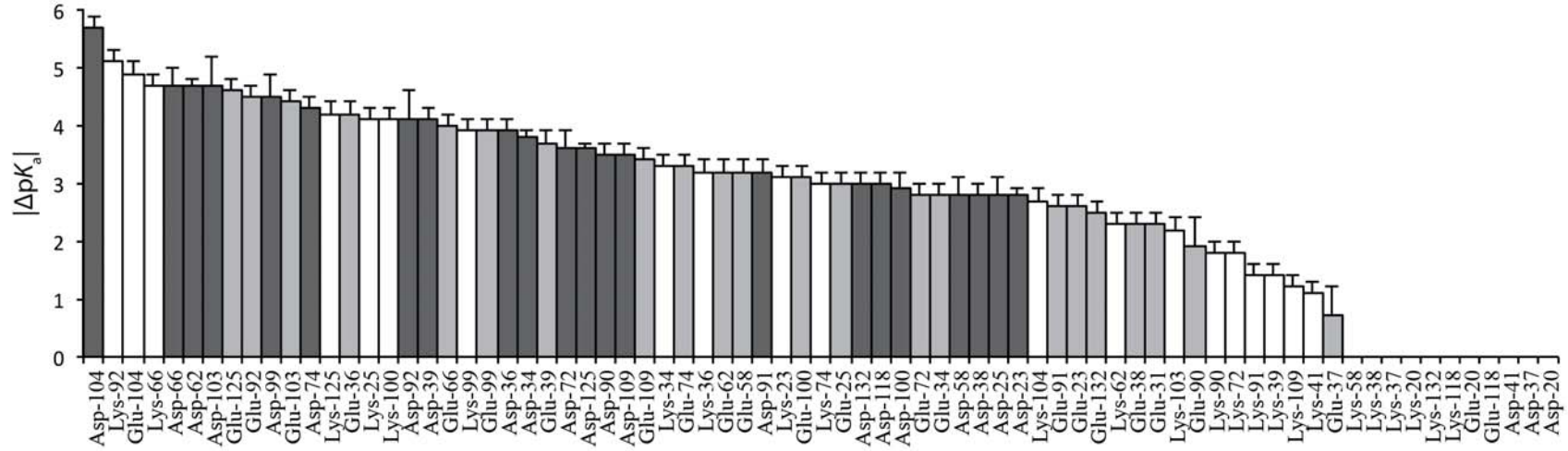

Figure 1.1 pK_a shifts for internal Asp, Glu, and Lys in SNase

Measured pK_a value shifts in rank-order for Asp (dark grey), Glu (light grey) and Lys (white). Error bars are the errors of the fit. All shifts are in the direction that favors the neutral state (elevated for basic Asp and Glu, depressed for Lys).

Of the 93 variants of SNase with internal Asp, Glu, Arg, or Lys residues that exhibited no large-scale structural rearrangement, 39 had local or subglobal structural changes of varying severity coupled with the ionization of the internal group observed by fluorescence and CD or NMR spectroscopy experiments. Some variants such as T62K[52,54] displayed highly localized fluctuations while others, such as L25K[52,53] and V104R[1], exhibited wide-spread structural perturbations. One variant, L38K, undergoes no discernable reorganization in the crystal structure, however an artificially high dielectric constant (ε_p ~30) is needed to reproduce the experimentally determined pK_a value implying that there is local conformational reorganization occurring[55]. The structural ensemble of the native state and other low-energy states must be incorporated into structure-based calculations for such algorithms to produce meaningful pK_a predictions.

Crystal structures have been determined for 40 variants of SNase with internal ionizable residues. In all but 8, the ionizable residues are deeply buried within the protein. In variants with substitutions at position 72 or with substitution to Arg, the side chain has limited but clear access to bulk water. Variants with substitutions at position 72 surprisingly have anomalous pK_a values even though the side chain can be interfacial with a favorable rotameric structure with no alteration of the global structure. The Arg variants all display some degree of structural reorganization that aids the establishment of hydrogen bonds to the guanadinium moiety to protein and water molecules. These variants are compelling examples that the complexities of the determinants of pK_a values. Further studies into how the structural dynamics of the protein modulate the pK_a values of internal ionizable residues are required.

1.2.3 Internal polar residues

In the attempt to discern some aspects of the shifted pK_a value of Asp and Glu residues, the thermodynamic properties of 48 variants with either an internal Asn or Gln substitution were examined and compared to the corresponding Asp and Glu variants[36]. Asn and Gln are the non-ionizable, isosteric counterparts to Asp and Glu. They are less well hydrated than Asp and Glu and are involved in fewer hydrogen bonds[56,57]. In general, at low pH where the pK_a of an internal Asp or Glu may not be shifted, Asp and Asn and Glu and Gln were quite similar. It was surprising that they were so similar given that their hydrogen bonding properties are distinctly different. In all cases, the substitution of an internal position with either Asn or Gln was destabilizing and comparable to the cost of burial of Lys in the neutral state.

1.3 Contributions from this dissertation

1.3.1 Thermodynamic consequences of substitution of internal positions with Ser, Thr, and Tyr

Chapter 2 of this dissertation describes the thermodynamic propterties of SNase variants in which 25 internal positions are substituted to Ser, Thr, and Tyr to assess the impact of spontaneous mutations to theses amino acids on thermodynamic stability and the ability of a protein to solvate these highly polar side chains. Substitutions to Ala and Val were also made at the same positions and compared to the variants with Ser, Thr, and Tyr to isolate the contributions related to the presence of a hydroxyl group. This was accomplished by measuring the thermodynamic stability by chemical denaturation monitored by intrinsic fluorescence and structural studies with circular dichroism

spectroscopy and x-ray crystallography. The data and insight garnered from these studies will be useful for the design and benchmarking of computational methods to examine consequences of mutations on protein stability as well as in the field of protein engineering.

1.3.2 Systematic survey of His residues buried in the protein interior

Chapter 3 presents a systematic survey of the properties of 25 internal His residues in SNase that complements previous studies with Asp, Glu, Lys, and Arg. The goal was to determine the extent to which the pK_a value of His is affected when buried in the protein interior. Histidine was of special interest because it has a unique chemical structure that makes is impossible to predict a priori if it would behave like Lys or like Arg. Accurate structure-based calculations of His pK_a values are extremely challenging and the intention of this study was to better understand properties such as hydrogen bonding and hydration by examining how His side chains behave in the microenvironments in the protein interior. The approach involved equilibrium thermodynamic measurements of protein stability by chemical denaturation monitored by fluorescence as well as spectroscopic and crystallographic experiments. These results will provide insight into important properties of His residue and help to improve existing methods for structure-based calculations that can more accurately predict pK_a values.

1.3.3 The pK_a value of an internal Lys residue is sensitive to the global thermodynamic stability of the protein

Previous work from this laboratory has shown that conformational reorganization coupled to the ionization of internal residues with anomalous pK_a values is an important determinant of these pK_a values[2,52,53,58]. The probability that partial unfolded states are populated is determined by the global thermodynamic stability of a protein ($\Delta G°_{H2O}$) because this measures the free energy difference between folded and partially unfolded states. If the pK_a value of an internal ionizable residue is determined by a population of partially or locally unfolded states where the ionizable residue makes contact with water, the pK_a should be sensitive to changes in the global stability of the protein. For example, the ionization of Glu at position 23 in a highly stable background of SNase stabilizes a partially unfolded conformation in which two of the beta strands in the OB domain is released[58]. Because the pK_a value is a population weighted average of the protein ensemble, the pK_a value of Glu-23 shifts from 7.5 to 7.1 in a slightly less stable form of SNase. Chapter 4 presents a study of the thermodynamic properties of Lys-23 and Ly-36 in 5 variants of SNase with $\Delta G°_{H2O}$ ranging from 8.0 to 13.8 kcal/mol. The pK_a values for Lys-23 and Lys-36 have previously been measured by means of thermodynamic linkage analysis as 7.3 and 7.2, respectively, in a SNase variant with $\Delta G°_{H2O}$ of 11.8 kcal/mol. The results are consistent with the proposed hypothesis and demonstrate that structure-based pK_a calculations cannot correctly predict the pK_a value for the proper physical reasons without the explicit treatments of high energy conformations that are populated in the conditions where the buried ionizable residue is exposed to water and becomes charged.

1.3.4 Use of internal ionizable groups with anomalous pK_a values to engineer pH sensitive switches active in the physiological range of pH

Chapters 5 and 6 present extensive experimental verification that ionizable residues with anomalous pK_a values can be used to engineer pH sensitive switches that are responsive to very small changes in pH within the physiological range. Nature tends to encode pH switching behavior in proteins through the use of multiple ionizable residues with altered pK_a values. This ensures resilience against spontaneous mutations that may remove key ionizable residues that contribute to the pH sensitivity. A rational and directed approach that attempts to reproduce these natural switches has turned out to be extremely difficult to achieve because the consequences of mutations that remove or alter ionizable groups can be unpredictable.

Other approaches towards the creation of a pH switch in the lab have relied on His residues and scanning libraries[59]. Our approach was founded on previously measured pK_a values of internal Lys and Glu residues[2,35] in SNase. To prove the concept, two positions in SNase were chosen to be mutated to either both Lys or Glu. We demonstrate that this can lead to a protein with stability that is acutely pH sensitive and unfolds near pH 7. The unfolding of the protein is entirely driven by the dramatic loss of stability as a function of pH resulting form the atypical pK_a values of the chosen residues. We also demonstrate that such switching behavior can be engineered with different pairs of internal Lys residues. The ability to regulate function, structure, and stability within the physiological pH range, as demonstrated in this thesis, creates many new opportunities in protein engineering and drug design.

Chapter 2

Thermodynamic consequences of burial of Ser, Thr, and Tyr residues in the interior of a protein

2.1 Abstract

The thermodynamic consequences of substitution of internal positions with the hydroxyl-bearing amino acids serine (Ser), threonine (Thr), and tyrosine (Tyr) were studied experimentally. This systematic study involved 75 variants of a highly stable form of staphylococcal nuclease with Ser, Thr, or Tyr at each of 25 internal positions. Thermodynamic stability measured by chemical denaturation monitored with Trp fluorescence showed that, on average, substitution of internal positions with Ser, Thr, or Tyr decreased the stability of the protein by 3.3, 3.2 and 4.5 kcal/mol, respectively. This is comparable to the cost of substitution of Lys, Asp, or Glu under conditions of pH where they are usually neutral. Comparison of the consequences of substitutions with Ser and Ala, or substitutions with Thr and Val, showed that on average the burial of a hydroxyl group destabilizes the protein by approximately 1.0 kcal/mol and the substitution of a methyl group with a hydroxyl by 1.2 kcal/mol. Structural consequences from the substitutions to Ser, Thr, or Tyr were largely undetectable in far UV-Vis spectra, consistent with what was observed with four representative crystal structures showing that the overall structure of the protein was not affected by the substitutions of hydrophobic to polar side chains. Only small local structural rearrangements of side chains near the site of substitution and the presence of internal water molecules were observed in the crystal structures. The destabilization of the protein by substitution of internal positions with Ser, Thr, or Tyr does not correlate with any structural metric, such as location, or identity of the original side chain.

2.2 Introduction

The availability of extensive databases with genomic information constitutes an opportunity to learn to interpret mutations in genes in terms of consequences to the structure, stability, and the function of the proteins they encode. To this end, quantitative, physical insight into structural and thermodynamic consequences of substitutions, especially of the non-conservative substitutions that are most likely to have unpredictable and significant consequences, will be invaluable. These non-conservative substitutions are not likely to be statistically well represented in sequence databases precisely because they can have highly deleterious consequences. Here we report on a systematic examination of the thermodynamic consequences of substitution of 25 internal positions in a globular protein with Ser, Thr, and Tyr.

Previously we demonstrated that internal positions in highly stable proteins can be substituted with Lys, Asp, and Glu[60]. These buried Lys, Asp, or Glu residues invariably lower the thermodynamic stability of the protein because the loss of hydration upon burial of the polar side chains is not compensated well by contact with polar moieties of the protein proper or with buried water molecules[2,60,61]. The side chains of Lys, Asp, and Glu are ionizable; therefore, the thermodynamic consequence of burial of these residues has two components. One is pH sensitive and is related to the shift in pK_a value experienced by the ionizable residue when buried in a hydrophobic environment[2,60,61]. This shift in pK_a promotes the neutral form of the ionizable group thereby minimizing the energetic penalty for removing a charge from water. The second component is independent of pH and is related to the removal of a polar side chain (i.e. Glu, Asp, or Lys side chains in the neutral state) from bulk water and burial in an

environment that is less polar and polarizable[2,61]. This component can be measured experimentally for Lys, Asp, and Glu by measuring thermodynamic stability under conditions of pH close to the normal pK_a of these side chains in water (pH 10 for Lys and pH 4 for Asp and Glu). We have also demonstrated that the energetic consequences of burial of the Arg side chain cannot be decomposed into pH dependent and pH independent components because the normal pK_a of Arg is close to 14 and the Arg side chain does not shift its pK_a nor is it ever found in the neutral state even when buried in hydrophobic environments[1,40,62].

To better understand the molecular determinants of the loss of stability related to burial of Lys, Asp and Glu side chains in the hydrophobic interior of a protein, we have now examined the cost of burying the polar residues Ser, Tyr, and Thr, all residues that have a side chain hydroxyl group. The polar side chains of Ser, Thr, and Tyr are usually found on the surface of proteins, where the polar moiety can interact with bulk water[3]. Polar residues that are partially or completely buried in the protein interior are not uncommon in proteins. They are usually involved in processes governed by electrostatic forces, such as catalysis[7,8], H^+ transport[18,19] and e^- transfer[16,17]. On the other hand, spontaneous mutations that result in the gain or loss of hydroxyl-containing amino acids, like Ser or Thr, have been associated with several diseases[63–65]. Presumably, in these cases, the substitution destabilizes the native state either because the removal of a polar group from water and burial in a hydrophobic environment is unfavorable or because removal of a side chain complementary to a polar pocket in the protein interior is unfavorable. Either way, destabilization leads to the loss of function either through unfolding or other structural reorganization in response to the presence of an

uncompensated hydrogen bonds in the hydrophobic core. To gain insight into how buried polar groups enable or disable biological function it is necessary to understand how they affect the thermodynamic stability of the folded state. That was the main goal of this study.

Previous studies have shown that buried polar groups are likely to destabilize the native state. For example, to increase the thermodynamic stability of a protein, apoflavodoxin, internal polar groups were replaced with non-polar ones[66]. Other studies have examined the extent to which intramolecular hydrogen bonding could compensate for the loss of contact between a buried polar group and water. These studies showed that buried polar groups can be extremely destabilizing – by as much as 3 kcal/mol – when their hydrogen bonding potential is not satisfied by either a water molecule or nearby protein atoms[5,6,67,68]. Because of the limited nature of these studies, they do not reveal the general trends in the magnitude of destabilization wrought by burial of polar groups in a hydrophobic environment or the molecular factors that determine the magnitude of this effect.

To address these issues, we have studied a family of 75 variants of a highly stable form of staphylococcal nuclease (SNase), engineered by introducing of Ser, Thr, or Tyr at each of 25 internal positions. The thermodynamic stability of each variant was measured at pH 7 by chemical denaturation monitored with Trp fluorescence. Crystal structures of four representative proteins were obtained to examine the microenvironments of these internal polar residues and structural consequences of the substitutions with Ser, Thr or Tyr. Far UV-CD was used to examine global consequences on three-dimensional structure. Variants with Ala or Val at the 25 internal positions were

also studied. By comparing variants with Ala or Ser, and variants with Val or Thr, it was possible to examine the energetic consequences of introduction of a single OH group in the protein interior and substitution of a CH_3 with an OH. The cost of substitution with Asn and Gln have also been measured[36].

Our results describe the range of destabilization associated with introduction of a single, uncompensated polar moiety in the protein interior. The data constitute an unprecedented set useful for the benchmarking of force fields and of computational methods for structure-based energy calculations. They will be especially useful for calibration of computational algorithms for studies of consequences of mutations on the properties of proteins.

2.3 Materials and methods

2.3.1 Proteins

The Stratagene Quickchange kit was used to engineer the variants, using the highly stable form of SNase known as Δ+PHS as the background protein. Each protein was expressed in *E. coli* BL21/DE3 cells (Invitrogen) transformed with the plasmid Pet24a+. Proteins were expressed and purified by the method of Shortle and Meeker[69] as modified by Byrne et al.[70]

2.3.2 Equilibrium thermodynamics

The Gibbs free energy of unfolding ($\Delta G°_{H2O}$) was measured using the intrinsic fluorescence of Trp-140 to monitor unfolding, as described previously[71]. GdmCl (UltraPure grade, Invitrogen Life Technologies) was used as a denaturant, as described

previously[71]. All measurements were performed with an ATF-105 automated fluorometer (Aviv Inc.) at 25℃. The protein concentration in these experiments was 50 μg/mL in a buffer consisting of 100 mM NaCl with 25 mM HEPES at pH 7.

2.3.3 CD spectroscopy

An Aviv model CD-215 circular dichroism spectrophotometer was used to collect far-UV CD spectra. Samples consisted of 50 μg/mL protein, 100 mM KCl and 25 mM TrisHCl. Spectra were collected for a volume of 1 mL of protein sample in a 0.1 cm path-length quartz cuvette at 25℃ at pH 7. Measurements were taken at wavelength intervals of 1 nm using an averaging time of 5 s.

2.3.4 X-ray crystallography

Variants were crystallized by the hanging drop vapor diffusion method at 4℃. The reservoir solution consisted of 20-45% (v/v) 2-methyl-2,4-pentanediol (MPD) and 15% glycerol in 25 mM potassium phosphate buffer (pH 6-9). Two molar equivalents of the inhibitor of 3'-5'-thymidine diphosphate (pdTp) and 3 molar equivalents of $CaCl_2$ were added to the protein solution before it was mixed with an equal volume of reservoir solution. The pdTp was synthesized in our laboratory[72].

Diffraction data were collected at 100K using a Bruker Duo Apex diffractometer with the exception of V23S, which was collected at Brookhaven National Labs using the NSLS X25 beamline. Initial phasing for all structures was obtained by maximum likelihood-based molecular replacement method with Phaser software within the CCP4 suite using a previously solved structure for Δ+PHS (PDB ID: 3BDC) as a search model.

Prior to molecular replacement, 3BDC.pdb was modified by truncating the substituted amino acid for the appropriate variant to Ala, removing all water molecules, and resetting all B-factors to 20.0 $Å^2$. Model building using Coot and refinement with Refmac5 were performed iteratively to yield the final models. R-work and R-free residuals were monitored throughout the refinement. Water molecules were added during model building to reflect spherical electron density in $2F_o$-F_c maps that were within 3.5 Å of a hydrogen bonding partner in the protein model. Final checks of the structures were done using SFCHECK and PROCHECK programs.

2.4 Results

2.4.1 Thermodynamic stability

The hyperstable variant of SNase, Δ+PHS, was used for thermodynamic stability measurements. It includes the G50F, V51N, P117G, H124L, and S128A substitutions, and a deletion of residues 44-49. The 25 positions that were substituted with Ser, Thr, or Tyr are listed in Table 2.1 and shown in Figure 2.1. The thermodynamic stability ($\Delta G°_{H2O}$) of each variant was measured by chemical denaturation with GdnHCl monitored by intrinsic Trp-fluorescence at pH 7 (Table 2.1, Fig. 2.2) as described previously[71]. The exception was the A132Y variant because the presence of Tyr-132 interfered with the fluorescence signal of Trp-140. For this variant, $\Delta G°_{H2O}$ was monitored by circular dichroism spectroscopy at 222 nm. Changes in stability relative to the parent protein ($\Delta\Delta G°_{H2O} = \Delta G°_{H2O,variant} - \Delta G°_{H2O,\Delta+PHS}$) show that the burial of a polar side chain lowered the

Table 2.1 Thermodynamic stability of SNase variants with Ser, Thr, and Tyr substitutions measured at 298 K in 100 mM ionic strength.

Site of modification[a]	Ser		Thr		Tyr	
	$\Delta G°_{H2O}$[b]	m-value[c]	$\Delta G°_{H2O}$	m-value	$\Delta G°_{H2O}$	m-value
Internal						
G20	11.5 (0.1)	4.7 (0.1)	9.3 (0.1)	5.0 (0.1)	9.1 (0.1)	5.0 (0.1)
V23	7.3 (0.1)	5.4 (0.1)	8.6 (0.2)	5.3 (0.1)	9.0 (0.1)	5.2 (0.1)
L25	7.1 (0.1)	5.6 (0.1)	11.5 (0.1)	4.8 (0.1)	6.2 (0.1)	5.3 (0.1)
F34	6.3 (0.1)	5.2 (0.1)	8.5 (0.1)	5.5 (0.1)	9.2 (0.1)	5.1 (0.1)
L36	6.7 (0.1)	5.8 (0.1)	8.0 (0.1)	5.4 (0.1)	9.2 (0.1)	5.4 (0.1)
L37	9.1 (0.1)	4.9 (0.1)	8.1 (0.2)	4.3 (0.1)	9.4 (0.1)	4.9 (0.1)
L38	10.3 (0.1)	4.8 (0.1)	9.0 (0.1)	5.1 (0.1)	9.9 (0.1)	5.0 (0.1)
V39	10.0 (0.1)	5.1 (0.1)	10.0 (0.1)	4.8 (0.1)	6.6 (0.1)	5.6 (0.1)
T41	11.1 (0.1)	4.9 (0.1)	-	-	9.5 (0.2)	5.1 (0.1)
A58	10.7 (0.1)	5.0 (0.1)	8.9 (0.1)	4.8 (0.1)	6.9 (0.1)	5.3 (0.1)
T62	9.4 (0.1)	4.9 (0.1)	-	-	9.0 (0.1)	5.2 (0.1)
V66	8.7 (0.2)	5.0 (0.1)	7.7 (0.2)	4.7 (0.1)	10.9 (0.2)	4.9 (0.1)
I72	6.3 (0.2)	5.4 (0.1)	7.8 (0.2)	5.0 (0.1)	7.5 (0.1)	5.1 (0.1)
V74	6.8 (0.2)	5.5 (0.1)	8.6 (0.3)	5.4 (0.2)	8.0 (0.1)	5.1 (0.1)
A90	9.7 (0.2)	5.0 (0.1)	10.3 (0.3)	5.1 (0.1)	6.1 (0.1)	5.4 (0.1)
Y91	6.6 (0.1)	5.7 (0.1)	7.0 (0.1)	5.1 (0.1)	-	-
I92	9.1 (0.1)	4.9 (0.1)	7.3 (0.3)	5.3 (0.2)	4.7 (0.1)	5.7 (0.1)
V99	6.1 (0.3)	4.2 (0.4)	11.0 (0.4)	5.1 (0.2)	5.0 (0.1)	5.7 (0.1)
N100	8.9 (0.1)	5.0 (0.1)	7.3 (0.1)	5.2 (0.1)	1.3 (0.1)	3.7 (0.1)
L103	7.2 (0.1)	4.8 (0.1)	7.7 (0.1)	5.3 (0.1)	5.5 (0.1)	5.8 (0.1)
V104	8.6 (0.1)	5.1 (0.1)	9.7 (0.1)	5.1 (0.1)	4.3 (0.1)	5.5 (0.1)
A109	10.5 (0.1)	5.0 (0.1)	9.6 (0.1)	5.2 (0.1)	8.1 (0.1)	5.2 (0.1)
N118	10.8 (0.1)	4.9 (0.1)	10.2 (0.1)	4.8 (0.1)	8.2 (0.1)	5.2 (0.1)
L125	6.6 (0.1)	5.5 (0.1)	7.3 (0.1)	5.3 (0.1)	10.3 (0.1)	4.9 (0.1)
A132	9.6 (0.1)	5.0 (0.1)	7.7 (0.1)	5.2 (0.1)	4.7 (0.2)	4.0 (0.2)
Average	8.6 ± 1.7	-	8.7 ± 1.3	-	7.4 ± 2.3	-
Surface						
A60	10.6 (0.1)	4.7 (0.1)	10.9 (0.1)	4.8 (0.1)	-	-
A69	10.1 (0.2)	4.9 (0.1)	9.0 (0.4)	4.9 (0.2)	-	-
A112	11.2 (0.2)	4.7 (0.1)	10.9 (0.2)	4.6 (0.1)	-	-
A128	10.4 (0.2)	4.9 (0.1)	9.8 (0.1)	5.1 (0.1)	-	-
Average	10.5 ± 0.5	-	10.2 ± 0.9	-	-	-

[a] All mutations were made in the Δ+PHS variant of SNase: $\Delta G°_{H2O}$ = 11.9 ± 0.1 kcal/mol[60], m-value = 4.9 ± 0.1 kcal/(mol * [GdnHCl])

[b] $\Delta G°_{H2O}$ in kcal/mol. Error of the fit is displayed in parenthesis.

[c] m-value in kcal/(mol * [GdnHCl]) from linear extrapolation.

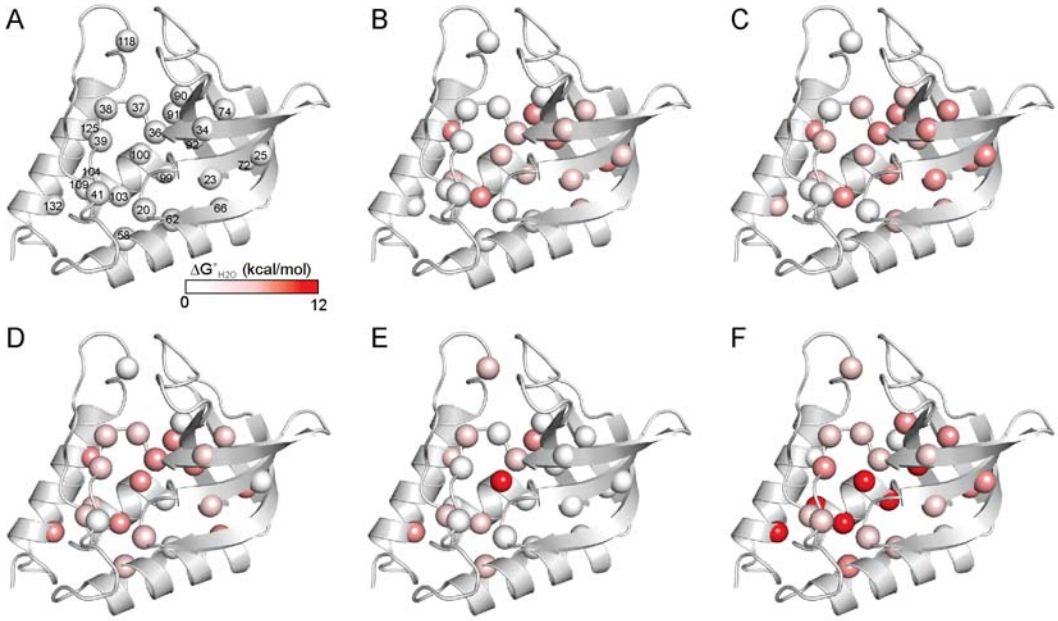

Figure 2.1 Structural distribution of thermodynamic stability represented on the structure of Δ+PHS. (A) Ribbon diagram of SNase (3BDC) with 25 internal positions represented as spheres. Temperature map of thermodynamic stabilities ($\Delta G°_{H2O}$) of variants with internal substitutions to **(B)** Ala, **(C)** Ser, **(D**) Thr, **(E)** Val, and **(F)** Tyr. Spheres representing substitution positions are colored based on $\Delta G°_{H2O}$ ranges: white is 12-10 kcal/mol, light pink is 10-8 kcal/mol, dark pink is 8-6 kcal/mol, red is 6-0 kcal/mol.

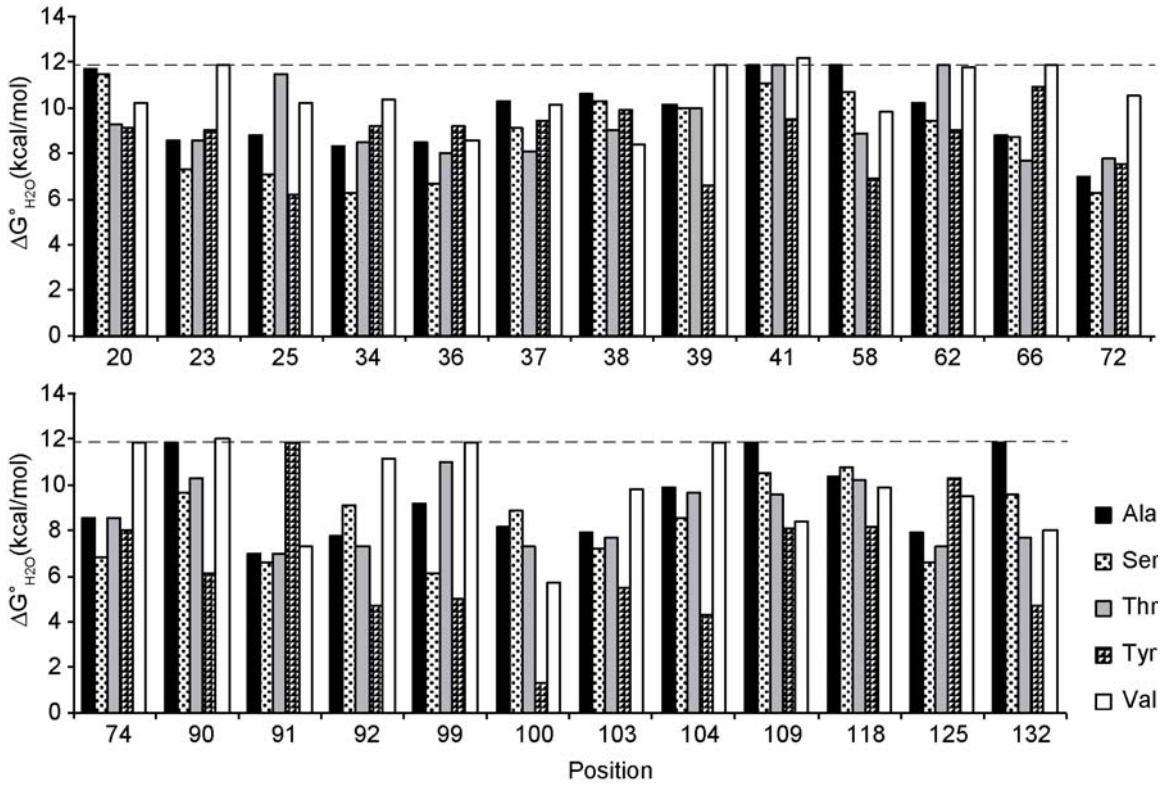

Figure 2.2 Thermodynamic stabilities ($\Delta G°_{H2O}$) of SNase variants with substitutions that put Ala, Ser, Thr, Tyr, or Val at each of 25 internal positions. Data are ordered be position number, and identified as Ala (black), Ser (dotted), Thr (grey), Tyr (checkered), and Val (white). The dashed line represents the thermodynamic stability of Δ+PHS at 298 K, 100 mM ionic strength.

stability of the native state by 0.4 to 5.8 kcal/mol (average $\Delta\Delta G°_{H2O,Ser}$ -3.3 ± 1.7 kcal/mol) for Ser substitutions, and by 0.4 to 4.9 kcal/mol (average $\Delta\Delta G°_{H2O,Thr}$ -3.2 ± 1.3 kcal/mol) for Thr substitutions (Table 2.1, Fig. 2.3). Substitution with Tyr was, on average, more destabilizing; $\Delta\Delta G°_{H2O}$ values ranged from -1.0 to -10.8 kcal/mol and $\Delta\Delta G°_{H2O,Tyr}$ = -4.3 ± 2.7 kcal/mol (Table 2.1, Fig. 2.3).

Four Ala residues on the protein surface were also substituted with both Ser and Thr to serve as controls. Even though these positions are all water-exposed, substitution to either Ser or Thr was destabilizing, on average by 1.3 ± 0.5 and 1.7 ± 0.9 kcal/mol, respectively (Table 2.1).

In attempts to isolate the contribution of the hydroxyl moiety to the destabilization of the native state, we compared the stability of variants with Ser and Thr substitutions to variants with Ala and Val substitutions, respectively. The thermodynamic consequences of the substitutions to Ala at the 21 internal positions that were not Ala in the reference protein were measured (Table 2.2, Fig. 2.4) as were the consequences of substitution with Val at the 19 that are not Val in the reference protein (Table 2.2). The Ala substitutions were destabilizing by 0 to 4.9 kcal/mol, with an average $\Delta\Delta G°_{H2O,Ala}$ value of -2.8 ± 1.4 kcal/mol (Table 2.2). The Val substitutions were destabilizing by -0.3 to 6.2 kcal/mol, with an average $\Delta\Delta G°_{H2O,Val}$ value of -2.2 ± 1.7 kcal/mol (Table 2.2). At positions 41 and 90 the substitutions with Val were not destabilizing ($\Delta\Delta G°_{H2O,Val}$ = 0.1 ± 0.1 kcal/mol for 41 and $\Delta\Delta G°_{H2O,Val}$ = 0.3 ± 0.1 kcal/mol for 90).

Table 2.2 Thermodynamic stability of SNase variants with Ala and Val substitutions measured at 298 K in 100 mM ionic strength.

	Ala		Val	
Site of modification[a]	$\Delta G°_{H2O}$[b]	m-value[c]	$\Delta G°_{H2O}$	m-value
G20	11.7 (0.4)	4.5 (0.1)	10.2 (0.1)	5.1 (0.1)
V23	8.6 (0.3)	4.9 (0.1)	-	-
L25	8.8 (0.2)	4.8 (0.1)	10.2 (0.2)	5.1 (0.1)
F34	8.3 (0.2)	5.1 (0.1)	10.4 (0.1)	5.0 (0.1)
L36	8.5 (0.2)	5.3 (0.1)	8.6 (0.2)	5.2 (0.1)
L37	10.3 (0.1)	4.7 (0.1)	10.1 (0.3)	4.6 (0.1)
L38	10.6 (0.2)	4.6 (0.1)	8.4 (0.1)	5.0 (0.1)
V39	10.1 (0.1)	5.0 (0.1)	-	-
T41	11.9 (0.2)	4.7 (0.1)	12.2 (0.2)	4.8 (0.1)
A58	-	-	9.8 (0.2)	5.0 (0.1)
T62	10.2 (0.1)	5.0 (0.1)	11.8 (0.1)	4.7 (0.1)
V66	8.8 (0.2)	4.7 (0.1)	-	-
I72	7.0 (0.2)	5.4 (0.2)	10.5 (0.1)	5.1 (0.1)
V74	8.6 (0.3)	5.1 (0.2)	-	-
A90	-	-	12.0 (0.1)	4.7 (0.1)
Y91	7.0 (0.1)	5.2 (0.1)	7.3 (0.1)	5.0 (0.1)
I92	7.8 (0.3)	5.2 (0.2)	11.2 (0.1)	4.9 (0.1)
V99	9.2 (0.7)	5.5 (0.4)	-	-
N100	8.2 (0.1)	4.9 (0.1)	5.7 (0.1)	4.9 (0.1)
L103	7.9 (0.1)	5.3 (0.1)	9.8 (0.1)	4.9 (0.1)
V104	9.9 (0.2)	4.9 (0.1)	-	-
A109	-	-	8.4 (0.1)	5.2 (0.1)
N118	10.4 (0.2)	4.9 (0.1)	9.9 (0.1)	4.9 (0.1)
L125	7.9 (0.1)	5.2 (0.1)	9.5 (0.1)	5.0 (0.1)
A132	-	-	8.0 (0.1)	5.2 (0.1)
Average	9.1 ± 1.4	-	9.7 ± 1.7	-

[a] All mutations were made in the Δ+PHS variant of SNase: $\Delta G°_{H2O}$ = 11.9±0.1 kcal/mol[60], m-value = 4.9 ± 0.1 kcal/(mol * [GdnHCl])
[b] $\Delta G°_{H2O}$ in kcal/mol. Error of the fit is displayed in parenthesis.
[c] m-value in kcal/(mol * [GdnHCl]) from linear extrapolation.

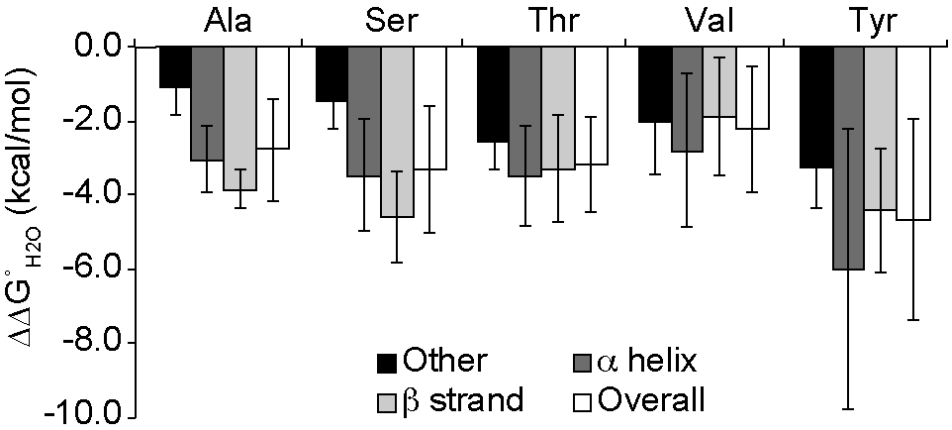

Figure 2.3 Average thermodynamic destabilization ($\Delta\Delta G°_{H2O} = \Delta G°_{H2O,variant} - \Delta G°_{H2O,\Delta+PHS}$) for substitution at 25 internal positions in SNase with Ala, Ser, Thr, Val, and Tyr. Substitution positions are binned according to secondary structural elements: overall (white), β-strand (light grey), helix (dark grey), other (black). Error bars report the standard deviation.

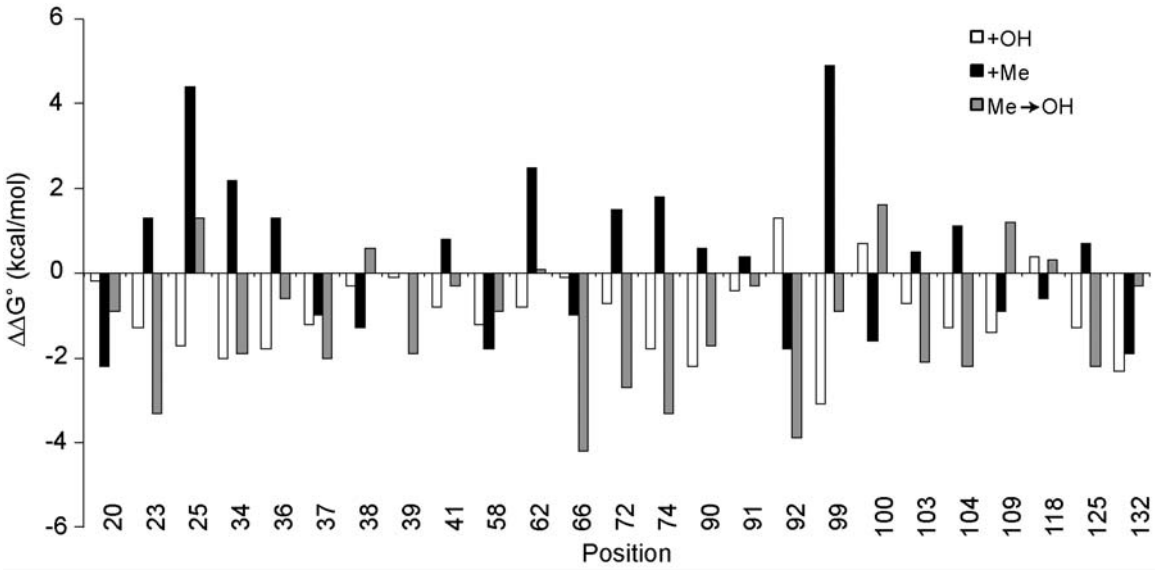

Figure 2.4 Contributions from methyl and hydroxyl groups to the thermodynamic stability of a protein. Thermodynamic consequences from addition of a methyl (black) ($\Delta G°_{Thr}$ - $\Delta G°_{Ser}$), from addition of a hydroxyl (white) ($\Delta G°_{Ser}$ - $\Delta G°_{Ala}$), and from substitution of a methyl for a hydroxyl (grey) ($\Delta G°_{Val}$ - $\Delta G°_{Thr}$) at 25 internal positions in SNase. A positive $\Delta\Delta G°$ value indicates an increase in thermodynamic stability; a negative value indicates the opposite..

2.4.2 CD spectroscopy

To examine how the global structure of the protein was affected by the substitution with Ser, Thr, Val, and Tyr, far-UV CD spectra were measured for the variants with internal Ser, Thr, Val, or Tyr substitution and compared to the parent protein (Fig. 2.5). Twenty-four of the 25 variants with any one of these four amino acids had spectra are comparable to the parent protein. These proteins appear to be fully folded. Only spectra of variants with substitutions at position 91, which is Tyr in the original protein, showed major deviations from that of the parent protein. In all three cases there was an increase in signal intensity at 208 nm accompanied by a decrease in signal 222 nm, which is reversed in the parent protein spectrum. This is not unique to substitutions with Ser, Thr, or Val, and has been observed in every other variant of Δ+PHS regardless of the amino acid substituted at this position[36,60,73].

2.4.3 Crystal structures

To establish that the hydroxyl groups are buried, to characterize the microenvironment of the polar side chains, and the potential of structural rearrangements associated with the substitutions, crystal structures for variants with substitutions at V23S, L25Y, I92S, and I92T were refined to a resolution of 1.6 Å, 2.1 Å, 1.8 Å and 1.9 Å respectively (Fig. 2.6, Fig. 2.7, Table A1). The crystal structures for all four variants were comparable to the structure of the parent protein (PDB ID: 3BDC), with Cα RMSD of 0.2 Å relative to 3BDC.

In the crystal structure of the V23S variant, the Ser side chain clearly resides inside the β-barrel in the same location as the original Val-23 side chain. The Oγ is fully

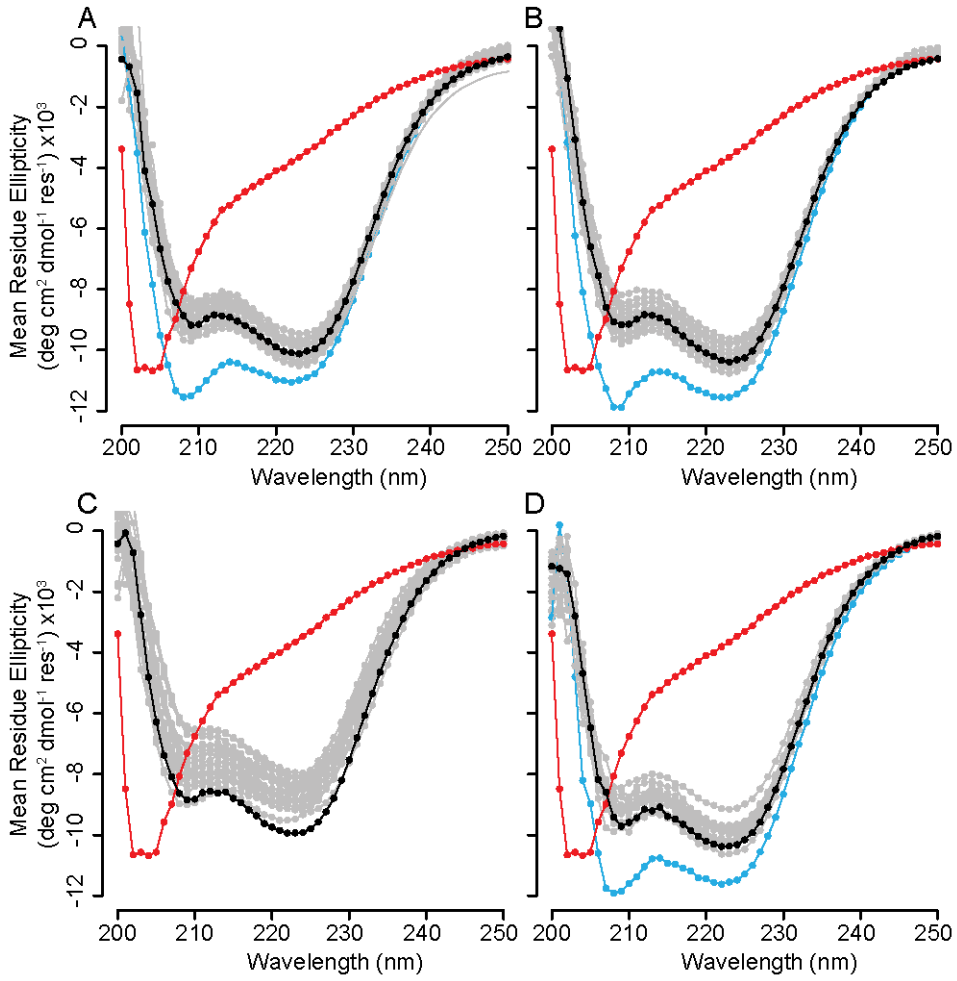

Figure 2.5 Far-UV CD spectroscopy of internal polar variants. Secondary structure measured by far-UV CD spectroscopy for variants of SNase with internal positions substituted with Ser (A), Thr (B), Tyr (C), and Val (D). Four types of CD spectra are shown in each panel: Δ+PHS SNase (black), unfolded T62 variant of the wild type (red), single substitutions (grey), variants with substitutions at Y91 (blue).

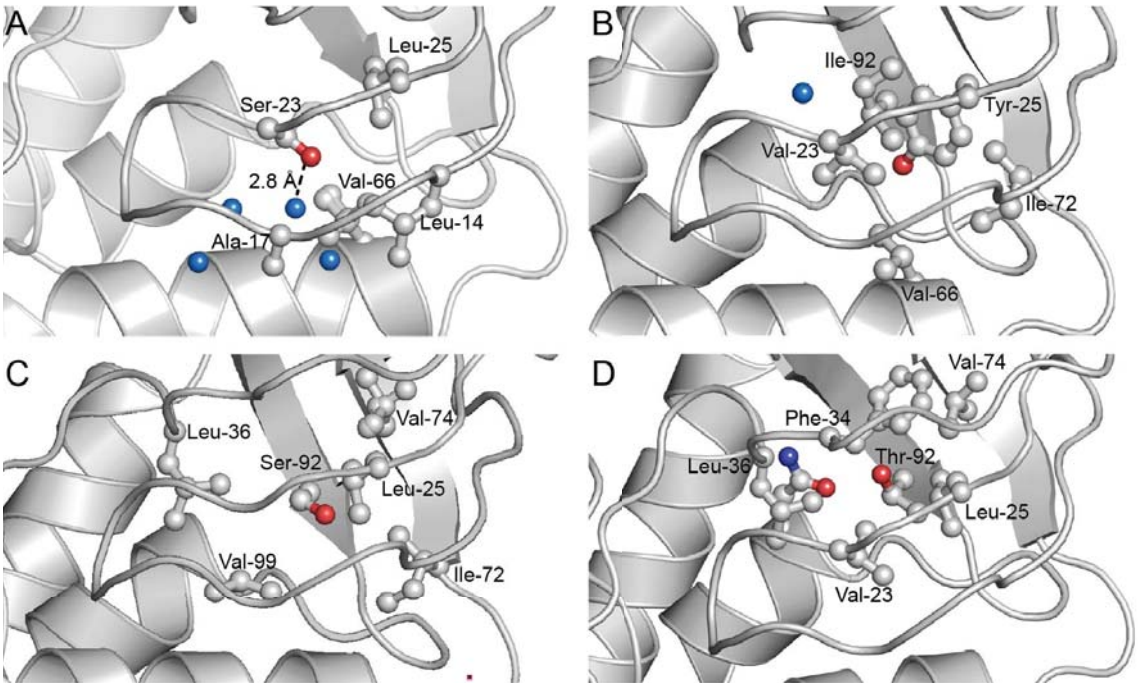

Figure 2.6 Microenvironments of internal Ser, Thr, and Tyr residues in SNase. Microenvironments of polar residues observed in variants with **(A)** V23S (PDB ID 4KHV), **(B)** L25Y, **(C)** I92S (PDB ID 4PMB) and **(D)** I92T (PDB ID 4ZQ3). The substituted residue is shown as ball and sticks with nearby waters represented as blue spheres and residues within 5 Å of the hydroxyl group of the substitution depicted as ball and sticks. For clarity, strands 1-3 were made transparent and are represented as coils.

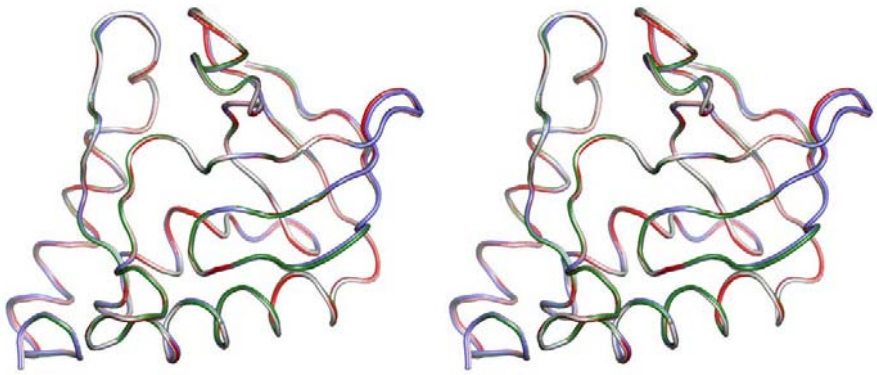

Figure 2.7 Global comparison of crystal structures containing internal polar mutations. Stereo image overlaying the backbone traces of crystallographically determined structures of Δ+PHS (white), Δ+PHS I92S (salmon), Δ+PHS V23S (green), Δ+PHS I92T (red) and Δ+PHS L25Y (blue).

sequestered from bulk solvent and resides in the microcavity in this region of SNase. The hydroxyl group points towards β-strand 1 and its sole potential hydrogen bonding partner is an internal water molecule 2.8 Å away, lodged between helix 1 and β-strand 1. This water molecule was not observed in the parent Δ+PHS structure, suggesting that it was stabilized in its position by a hydrogen bond with Ser-23.

In the crystal structure of the variant with the L25Y substitution the side chain of the Tyr-25 was fully internal. Its hydroxyl group points towards helix 1. A water molecule that was not detected in the structure of the parent protein was observed inside the β-barrel near β-strands 2 and 3. However, this water molecule was 5.3 Å away from the hydroxyl oxygen, too far to hydrogen bond with the Tyr. There are no potential hydrogen bonding partners within 5 Å of the hydroxyl group on the Tyr.

The side chain of Ser-92 is fully internal and overlays well with the Cβ and Cγ2 atoms of Ile-92 in the structure of the parent protein, Δ+PHS. There are no potential hydrogen bonding partners present within 5 Å of the Oγ of Ser-92. Similarly, the side chain of Thr-92 is internal and overlays well with the Cβ, Cγ1, and Cγ2 atoms of Ile-92 in Δ+PHS. There are no visible water molecules in this crystal structure with no other potential hydrogen bonding partners within 5 Å of the Oγ1 atom on Thr-92.

2.5 DISCUSSION

2.5.1 Variability in response at 25 internal positions

The virtue of studying thermodynamic consequences of substitutions systematically, at 25 internal positions in the same protein, is that it reveals the wide variability of response to the substitutions, the difficulty in correlating observed behavior

with any available structural metric, and the perils of over-interpreting measurements of single positions or average values. This systematic study reveals clearly that even in the same protein, at two deeply buried positions, substitutions of the same type can have very different consequences (i.e. V66Y and I92Y). Even in cases where contributions from the unfolded state can be eliminated by comparing only the substitutions at positions that have the same residue in the parent protein (i.e. L25, L36, L37, L38, L103, L125), substitution to the same type of residue (e.g., Ser) can have very different consequences. The thermodynamic consequences of the substitutions at the 25 internal positions are governed by a complex mixture of factors: differences in side chain conformational entropy, hydrogen bonding patterns, van der Waals interactions and packing, and the extent to which water penetration is promoted. Even with this extensive data set it is impossible to parse out these effects. Future computational studies should focus on reproducing the distribution of response illustrated by the data in Tables 2.1 and 2.2.

2.5.2 Differential properties of surface and internal positions

Since polar moieties are more favorable in polar environments such as water, mutations that introduce Ser, Thr, or Tyr at the protein-water interface are expected to affect the stability of the folded state less than mutations that introduce them in the interior. Previous studies[68,70] have examined mutations from surface nonpolar residues to Ser or Thr and reported $\Delta\Delta G°_{H2O}$ values ranging from +0.2 to -4.8 kcal/mol with an approximate average of -2 kcal/mol. The mutations reported here at surface positions where an Ala residue was substituted with a Ser or Thr residue have an average destabilizing effect of 1.3 kcal/mol and 1.7 kcal/mol, respectively, which is well within

the scope of previously reported values. The variation in the magnitude of protein destabilization from even these seemingly small substitutions indicates that even if the hydroxyl group is well hydrated, other factors such as loss of side chain entropy, differences in hydrogen bonding interactions, and perturbation to van der Waals interactions, can make a significant difference. The four Ala→Ser and Ala→Thr substitutions at internal positions are 0.5 kcal/mol and 1.1 kcal/mol more destabilizing than equivalent substitutions at surface positions. This is not surprising as burial of the side chain will decrease side chain conformational entropy and hydrogen bonding. The variation between different types of substitutions at the internal sites is greater than these differences between surface and internal Ala→Ser and Ala→Thr substitutions.

2.5.3 Comparison with internal, neutral Lys and Glu

The average cost of substituting these 25 internal positions with Lys at high pH where the Lys residues in water are neutral is -4.1 kcal/mol. This is the cost of substituting the internal positions with neutral Lys. This value represents the global thermodynamic consequence of removal of a Lys side chain from water and burial into the protein interior. Similarly, the cost of burying Glu at low pH where it is normally in the neutral state, is -3.8 kcal/mol. These value can now be compared with the thermodynamic cost of substituting the 25 internal positions at pH 7 with Tyr, Ser, and Thr, which were -4.1 ± 1.9 kcal/mol, -3.3 ± 1.7 kcal/mol and -3.2 ± 1.3 kcal/mol, respectively (Fig. 8). Although the error bars are large because the spread of values is also large, it would appear that the smaller side chains Ser and Thr can be accommodated

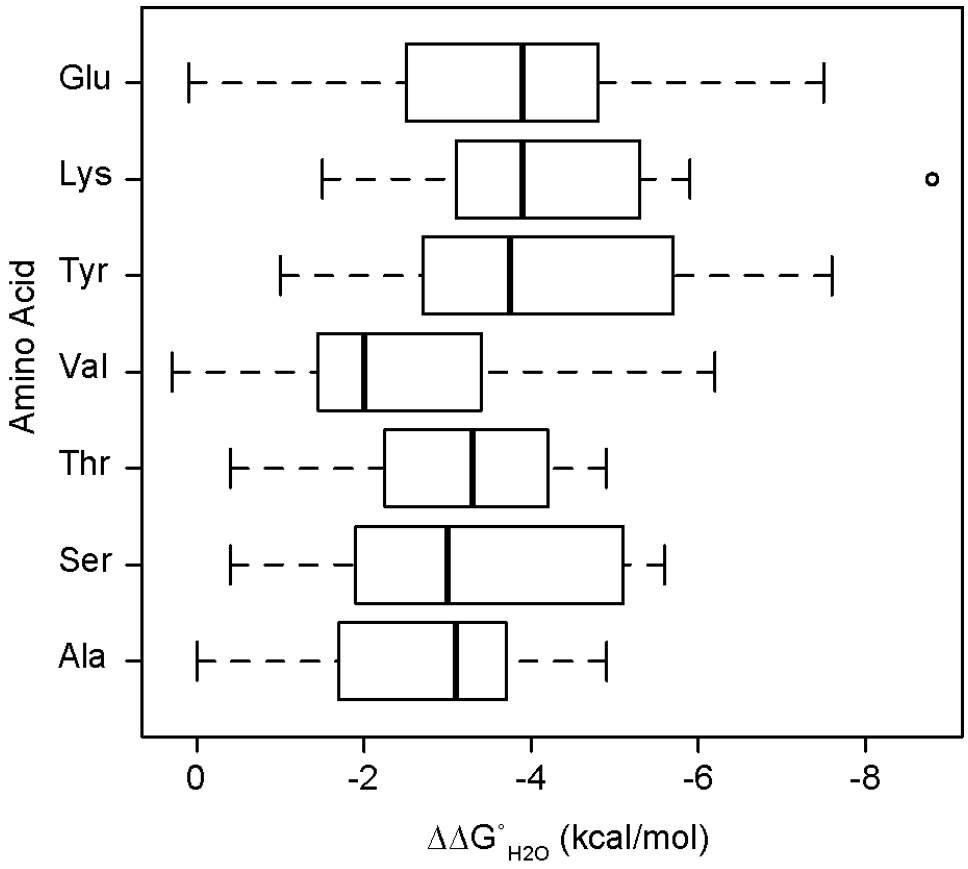

Figure 2.8 Comparison of $\Delta\Delta G°_{H2O}$ of polar and neutral ionizable substitutions. Distribution of thermodynamic effects of substitutions to polar or ionizable residues at 25 internal positions in SNase. All values were recorded at pH 7 except Lys (pH 10) and Glu (pH 4). Outlier is represented as circles and the median is indicated by a thick black line.

more easily in the hydrophobic interior of SNase that the bulkier Glu, Lys, or Tyr side chains. On the other hand, the destabilizing consequences of substitution of internal positions with Lys and Tyr appear to be comparable in this respect, and these side chains have very different properties.

2.5.4 Structural correlates of destabilization

In the case of variants with internal ionizable groups, the shift in pK_a values can be used as a diagnostic for burial in the hydrophobic interior of the protein. In several dozen SNase variants with internal Lys, Asp, and Glu, the expectation that the side chain of the residue introduced at an internal position is buried has been confirmed by crystal structures. In the case of these variants with internal Ser, Thr, or Tyr, the four structures that were determined confirm that, as expected, the side chains of Ser-23, Tyr-25, Ser-92 and Thr-92 are internal. On the other hand, these structures have little information about the molecular determinants of the thermodynamic consequence of substitution of introduction of polar residues by substitution at these internal positions. The structure with Ser-23 suggests that water penetration can ameliorate the thermodynamic consequence of burying a polar group, but internal water molecules were observed only in the structure of the variant of Ser-23, and the presence or absence of internal water molecules does not track with the magnitude of the effects of the V23S substitution on stability.

It is well known that the 20 amino acids differ in their propensities to be found in the different elements of secondary structure. In a recent statistical analysis of structures in the protein data bank[74], Tyr was shown to strongly favor β-strand conformations while

Ser favors random coil conformations. Thr appears to have an equal preference for β-strand and random coil elements. In the context of mutations to all polar side chains, some studies have seen no clear patterns between change in $\Delta G°_{H2O}$ and placement of the substitutions on different elements of secondary structure[70], while others saw that mutations with an increased hydrophilic nature were less destabilizing in coil regions than α-helix or β-strand regions[75].

The $\Delta\Delta G°_{H2O}$ values for the variants with internal Ser, Thr, and Tyr were organized according to the type of elements of secondary structure where the substitution was made (Fig. 2.3). The internal positions 20, 37-41, and 109-118 correspond to coil, positions 58-66, 99-104, and 125-132 are helical, and 23-26 and 72-92 are β-strand. As expected, Ser was less destabilizing in coil regions, where its destabilizing effects were small compared to what they were in both α-helix or β-strand regions. In contrast, neither Thr nor Tyr showed a significant difference between the three structural regions considered, only a slight, roughly 1 kcal/mol difference in average $\Delta G°_{H2O}$ favoring random coil. It is likely that the large size and aromatic nature of the Tyr side chain and the β-branched and more hydrophobic nature of the Thr side chain are complicating factors in this analysis.

The positions surrounding the β-barrel that define the hydrophobic core of SNase were amongst the ones that were most destabilizing when mutated to Ser, Thr, or Tyr. This cluster of positions was identified by Shortle et al. as corresponding to the main hydrophobic core of SNase[76]. This analysis involved the comparison of normalized m-values in single Ala and Gly variants. By applying the same analysis and classification for m^- (≤ 0.95), m^0, and m^+ (≥ 1.05) mutants to the set of Ser, Thr, and Tyr variants

(Table 2.1), the m^+ mutants clustered in the same positions in the β-barrel region identified previously by Shortle. This observation is in line with what has previously been reported[76,77]. All this suggests that, despite the more strongly destabilizing nature of substitutions of internal positions, the overall response of the protein follows the same pattern observed previously in saturation mutagenesis studies with Ala and Gly.

2.5.5 Thermodynamic consequences of burying an OH group

To determine the thermodynamic consequences of introducing a hydroxyl group in the relatively hydrophobic interior of SNase, $\Delta\Delta G_{OH} = (\Delta G_{Ser} - \Delta G_{Ala})$ was calculated at each position. The addition of a single hydroxyl group is, on average, destabilizing by 1.0 ± 1.0 kcal/mol (Table 2.3), with three notable exceptions. Positions 92, 100 and 118 are all more tolerant of a Ser than of an Ala. It is important to note here that positions 100 and 118 are both Asn in the unaltered sequence and participate in hydrogen bonds. Position 92 will be discussed below in detail.

A similar analysis was performed to determine both the cost of adding a methyl group ($\Delta\Delta G_{Me} = \Delta G_{Thr} - \Delta G_{Ser}$) and the substitution of a methyl for a hydroxyl ($\Delta\Delta G_{Me\rightarrow OH} = \Delta G_{Thr} - \Delta G_{Val}$) (Table 2.3). Overall, the addition of a methyl group was stabilizing by 0.4 ± 1.9 kcal/mol. In the β-barrel, which is well packed and where an increase in van der Waals interactions was expected, the addition of the methyl group increased the stability by 1.3 ± 1.6 kcal/mol. In coiled regions, which are heavily influenced by conformational entropy[75], the effect was destabilizing on average by -0.7 ± 0.9 kcal/mol.

Table 2.3 Thermodynamic consequences of hydroxyl burial, methyl incorporation, or methyl-hydroxyl substitution.

Position	Ser - Ala $\Delta\Delta G°_{H2O}$	Thr - Ser $\Delta\Delta G°_{H2O}$	Thr – Val $\Delta\Delta G°_{H2O}$
20	-0.2 (0.4)	-2.2 (0.1)	-0.9 (0.1)
23	-1.3 (0.2)	1.3 (0.2)	-3.3 (0.2)
25	-1.7 (0.2)	4.4 (0.2)	1.3 (0.1)
34	-2.0 (0.2)	2.2 (0.1)	-1.9 (0.1)
36	-1.8 (0.2)	1.3 (0.2)	-0.6 (0.1)
37	-1.2 (0.1)	-1.0 (0.5)	-2.0 (0.1)
38	-0.3 (0.2)	-1.3 (0.1)	0.6 (0.1)
39	-0.1 (0.1)	0.0 (0.1)	-1.9 (0.1)
41	-0.8 (0.2)	0.8 (0.2)	-0.3 (0.1)
58	-1.2 (0.1)	-1.8 (0.2)	-0.9 (0.1)
62	-0.8 (0.1)	2.5 (0.1)	0.1 (0.1)
66	-0.1 (0.3)	-1.0 (0.2)	-4.2 (0.1)
72	-0.7 (0.3)	1.5 (0.2)	-2.7 (0.1)
74	-1.8 (0.4)	1.8 (0.3)	-3.3 (0.1)
90	-2.2 (0.2)	0.6 (0.3)	-1.7 (0.1)
91	-0.4 (0.1)	0.4 (0.1)	-0.3 (0.1)
92	1.3 (0.3)	-1.8 (0.3)	-3.9 (0.1)
99	-3.1 (0.8)	4.7 (0.54	-0.9 (0.1)
100	0.7 (0.1)	-1.6 (0.1)	1.6 (0.1)
103	-0.7 (0.1)	0.5 (0.1)	-2.1 (0.1)
104	-1.3 (0.1)	1.1 (0.1)	-2.2 (0.1)
109	-1.4 (0.1)	-0.9 (0.1)	1.2 (0.1)
118	0.4 (0.2)	-0.6 (0.1)	0.3 (0.1)
125	-1.3 (0.1)	0.7 (0.1)	-2.2 (0.1)
132	-2.3 (0.1)	-1.9 (0.1)	-0.3 (0.1)
Avg.	-1.0 ± 1.0	0.4 ± 1.9	-1.2 ± 1.6

The difference in stabilities of Thr and Val mutations ($\Delta\Delta G_{Me \rightarrow OH}$) is destabilizing by 1.2 ± 1.6 kcal/mol and agrees with most of the $\Delta\Delta G_{OH}$ values in a positional comparison. Positions 100 and 118 are more stable with Thr than with Val, which reinforces the suggestion that positions that have been evolutionarily selected for polar amino acids are stabilized by the presence of hydroxyl groups[68].

Although the addition of new atoms when a hydroxyl group is introduced or when a methyl group is substituted for a hydroxyl are by and large comparable at most positions, some positions are sensitive to small differences. For example, position 25 is destabilized by 1.7 kcal/mol when a hydroxyl group is added yet is stabilized by 1.3 kcal/mol when a methyl group is replaced by a hydroxyl. A similar pattern is seen with position 109. However, position 92 shows the opposite behavior. The protein is stabilized by 1.3 kcal/mol when a hydroxyl group is added at this position and it is destabilized by 3.9 kcal/mol when a methyl was switched for a hydroxyl group. Based on the crystallographic data for position 92, the side chain of the Ser can potentially make a hydrogen bond with a nearby backbone carbonyl while the side chain of the Thr has no potential hydrogen bonding partners available to form a similar interaction. These results agree with previous studies suggesting that hydrogen bonding in an internal position can be sufficient to compensate for the destabilizing effect of side chain dehydration[5,6,78].

2.6 Conclusions

Substitution of internal, nonpolar residues in a protein with Ser, Thr, or Tyr destabilized the native state. The average magnitude of the effect is comparable to what was observed previously when buried Lys or Glu were introduced in the neutral states.

No distinguishable pattern can be detected that suggests that some polar moieties are easier or harder to accommodate in the protein interior. The consequences of the effects of the different types of polar side chains is different from position to position, suggesting that the magnitude of the effects are governed by a variety of factors that must include side chain entropy effects, van der Waals interactions and hydrogen bonding.

Despite the lack of simple structural or thermodynamic rules to interpret the data that were measured, the data set does contribute significant insight into the magnitude of nonconservative mutations that incorporate polar residues in the hydrophobic interior of the protein. The effects are not as dramatic as what has been observed when ionizable groups are buried in the hydrophobic interior of the protein, but they are significant enough to warrant special consideration in computational algorithms designed to explain potential consequences of mutations on structure, stability, and function of proteins.

Chapter 3

Properties of histidine side chains buried in the hydrophobic interior of a protein

3.1 Abstract

Ionizable residues buried in the hydrophobic interior of proteins are essential for function. Their properties can be highly anomalous and they are poorly understood. The pK_a value of Lys residues buried in 25 internal locations in staphylococcal nuclease (SNase) were measured previously and some were shown to be as low as 6. In contrast the pK_a values of Arg side chains buried at these same internal locations were unaffected; Arg remains charged over a wide range of pH values even if buried in hydrophobic environments. Given the asymmetry, aromaticity, delocalized charge, and multiple tautomeric states of the imidazole group, His could in principle behave in a dehydrated environment more like Arg than like Lys. This was examined by measuring the pH dependence of the thermodynamic stability of variants of SNase with His residues buried at 25 internal locations. These data showed that the pK_a values of buried His residues are sensitive to their microenvironment and that they are depressed. Crystal structures of 8 variants showed that the His side chains are buried, that they can coexist with a strictly hydrophobic microenvironment, but that they will hydrogen bond when this is a possibility. The data suggest that His residues buried in dehydrated environments in proteins behave more like Lys than like Arg. Studies are on going to examine what appears to be extreme sensitivity of the His side chain to the local polarity of its microenvironment.

3.2 Introduction

Ionizable residues buried in the hydrophobic interior of proteins can have highly anomalous and unpredictable properties that are essential for their biological functions.

Lysine residues sequestered in the protein interior can have depressed pK_a values that allow them to exist primarily in the neutral state near physiological pH and to participate as H^+ donors/acceptors in enzymatic catalysis[2,60,79]. The pK_a values of buried Asp and Glu residues shift in the opposite direction and can be as high as 10[60,61]. These anomalously high pK_a values are necessary for H^+ pumping in proteins such as bacteriorhodopsin[80,81], cytochrome c oxidase[20,82], and ATP synthase[19,83]. Arg is very different from Lys, Asp, and Glu. It retains its charge even when buried in the hydrophobic environment inside a membrane bilayer[62] or in the interior protein[1]. Arg is used in nature when a residue must remain charged even when removed from water, for example in bacteriorhodopsin[84], ATP synthase[85], and the voltage sensing paddle in the K^+ channel[62].

The properties of histidine residues when buried in the hydrophobic interior of proteins are not known. Histidine is of special interest because it is required for many biological processes, in coordination of ligands, metals and cofactors[22–27], and in catalysis[7,86,87]. Histidine is involved in the pH sensing mechanisms, for example, in the regulation of opening and closing the M2 proton channel in the influenza A virus[31,88], of the oxygen binding properties of hemoglobin[26], and in the activation of the influenza virus[89].

The versatile His residue is of special interest to the biomedical field as it is the ionizable residue that usually has a pK_a value within the physiological pH range. It participates in a wide variety of biological processes because of its rich chemistry. Besides being both polar and aromatic, it can participate in ion pairs[8], π-electron interactions[25], and as an acceptor or donor of H bonds[90]. Owing to the asymmetry of the

imidazole side chain the two nitrogen atoms have different proton affinities[33,91]; the observed pK_a of the His side chain can involve contributions from two different tautomeric states[43].

In contrast to Lys, Arg, Asp, and Glu, which are found predominantly at the protein-water interface where they are hydrated by bulk water, His residues are often sequestered from bulk water[92]. It is not known whether in its response to burial in the protein interior, His behaves like Arg, which can be buried in hydrophobic environments while fully charged, or like Lys, Asp, and Glu, which acquire anomalous pK_a such that they are buried in the neutral state to minimize the energetic penalty from dehydration. In principle, the delocalization of charge in the aromatic imidazole ring, van der Waals, π-electron interactions, and its ability to act as either hydrogen bond donor or acceptor could compensate for loss of contact with water. The matter is further complicated by the intrinsic difficulties with structure-based calculations of histidine pK_a values. These calculations are not yet sufficiently accurate to allow a meaningful dissection of the different factors that could affect the pK_a values of histidine residues in the complex microenvironments in the protein interior[93,94].

The histidine side chain has received considerable attention. NMR spectroscopy has been used for accurate and precise measurement of pK_a values[23,95,96]. Surveys of His pK_a values do not reveal general trends, especially concerning the properties of histidine residues buried in the protein interior[90,97]. It is well understood that the pK_a of His is affected by anything that affects the contributions from its different tautomeric states, but the factors that affect the tautomeric states are themselves not well understood[33,43].

We have now examined properties of 25 internal His residues in a highly stable

form of staphylococcal nuclease (SNase). These His residues were buried in SNase by mutagenic substitution at internal positions. The properties of Lys, Glu, and Arg buried at these 25 positions were studied previously[1,2,61]. The thermodynamic stability of the 25 His containing variants was measured by chemical denaturation at pH 5 and 7 to determine if the pK_a of the His was depressed relative to its normal value in water. Circular dichroism and NMR spectroscopy were used to identify any structural perturbations related to substitution of an internal position with His, and in some cases, the tautomeric and protonation state of the internal His was determined. Crystal structures were determined for some variants to examine the microenvironments of the buried His side chain. The acid unfolding of the proteins precluded determination of the pK_a of the buried His residues but the data demonstrate that the pK_a values of His residues are depressed when their side chains are buried in a hydrophobic environment. The results contribute significant insight into the role of hydration as an important determinant of the pK_a of the His residue. The data set that emerged from this study should serve as an essential benchmark for calibration of structure-based pK_a calculations.

3.3 Materials and methods

3.3.1 Proteins

The Stratagene Quickchange kit was used to engineer the variants, using the highly stable form of SNase known as Δ+PHS as the background protein. Each protein was expressed in *E. coli* BL21/DE3 cells (Invitrogen) transformed with the plasmid Pet24a+. Proteins were expressed and purified by the method of Shortle and Meeker[69] as modified by Byrne et al.[70]

3.3.2 Measurement of ΔG^{o} by chemical denaturation

The Gibbs free energy of unfolding ($\Delta G°_{H2O}$) was measured using the intrinsic fluorescence of Trp-140 to monitor unfolding, as described previously[71]. The far-UV signal at 222 nm was used for His-132 variant. GdmCl (UltraPure grade, Invitrogen Life Technologies) was used as a denaturant, as described previously[71]. All measurements were performed with an ATF-105 automated fluorometer (Aviv Inc.) at 25℃. The protein concentration in these experiments was 50 μg/mL in a buffer consisting of 100 mM NaCl with 25 mM HEPES at pH 7 or 25 mM MES at pH 5.

3.3.3 Far-UV CD Spectra

An Aviv model CD-420 circular dichroism spectrophotometer was used to collect far-UV CD spectra. Samples consisted of 50 μg/mL protein, 100 mM KCl and either 25 mM TrisHCl. (pH7) or 25 mM KAc (pH5). Spectra were collected for a volume of 1 mL of protein sample in a 0.1 cm path-length quartz cuvette at 25 ℃. Measurements were taken at wavelength intervals of 1 nm using an averaging time of 3 s.

3.3.4 NMR spectroscopy.

^{15}N-labeled protein samples in deionized, distilled water was exchanged into 100% D_2O buffer containing 25 mM sodium phosphate, 100 mM sodium chloride in Centricon YM-10 (Millipore, Billerica, MA) filter concentrators. Samples were diluted in buffer to yield a final concentration of 1 mM. The pH was measured for the sample (without correction for isotope effects) with an Orion Research model 720A pH meter using a 3 mm glass combination electrode (Mettler Toledo, Columbus, OH). Adjustment

of the sample pH was done using 1μL aliquots of concentrated DCl or NaOD at 5 ℃. Data were collected at 298 K on a Bruker Avance II-600 equipped with a cryoprobe. Initial conversion and processing of the data was done using NMRPipe[98] and analyzed with Sparky[99] .

3.3.5 X-ray crystallography.

The proteins were crystallized using the hanging drop vapor diffusion method. The reservoir solution contained 2-methyl-2,4-pentanedio (MPD) (Sigma-Aldrich Corp.) and 25 mM potassium phosphate. Proteins were mixed in a 1:1 ratio with reservoir solutions before suspension and incubation at 4℃. For the V23H, L25H, L36H, T62H, V66H, I92H variants, the protein was pre-incubated with 3'-5'-thymidine diphosphate (pdTp) and calcium chloride in a 1:2:3 ratio before mixing with the reservoir solution. For the I72H and Y91H variants, the ratio used was 1:1:2.

Diffraction data were collected at conditions describe in Table A2. Initial phasing for all structures was obtained by maximum likelihood-based molecular replacement method with Phaser software within the CCP4 suite using a previously solved structure for Δ+PHS (PDB ID: 3BDC) as a search model. Prior to molecular replacement, 3BDC.pdb was modified by truncating the substituted amino acid for the appropriate variant to Ala, removing all water molecules, and resetting all B-factors to 20.0 $Å^2$. Model building using Coot and refinement with Refmac5 were performed iteratively to yield the final models. R-work and R-free residuals were monitored throughout the refinement. Water molecules were added during model building to reflect spherical electron density in $2F_o$-F_c maps that were within 3.5 Å of a hydrogen bonding partner in

the protein model. Final checks of the structures were done using SFCHECK and PROCHECK programs.

3.4 Results

3.4.1 Thermodynamic stability

The thermodynamic stability, $\Delta G°_{H2O,}$ for the 25 His-containing variants was determined at pH 7 and pH 5 by chemical denaturation with guanidinium hydrochloride (Table 3.1 and Fig. 3.1). Titration curves for 24 of the 25 variants were obtained by monitoring by intrinsic fluorescence of Trp-140, which has shown to be a robust reporter of global structure of SNase[49,100,101]. In the case of His-132, the proximity of the residue to Trp-140 in the native conformation interfered with the fluorescent signal. For this variant, the far-UV CD signal at 222 nm was used monitor unfolding instead. All unfolding curves of the 25 variants displayed apparent two-state behavior and were analyzed accordingly with a two-state linear extrapolation model[102].

The consequences of His substitutions on thermodynamic stability at each pH was obtained by subtracting the $\Delta G°_{H2O}$ of the parent protein from the $\Delta G°_{H2O}$ of the His-containing variant: $\Delta\Delta G°_{H2O} = \Delta G°_{H2O,His} - \Delta G°_{H2O,\Delta+PHS}$ (Table 3.1). Under all conditions studied, the His-containing variants are less stable than the parent protein. At pH 7, where the His is assumed to be neutral, the $\Delta\Delta G°_{H2O}$ ranged from -1.2 to -7.3 kcal/mol with an average $\Delta\Delta G°_{H2O}$ of -4.7 ± 1.8 kcal/mol. There are no obvious structural metrics that correlate with the effects of the substitutions with His on thermodynamic stability. It does not correlate with what the original amino acid was at the internal location or with the

Table 3.1 Thermodynamic stability of variants of SNase with His residues at internal positions. All variants were made using the Δ+PHS form of SNase as background

	$\Delta G°_{H2O}$ (kcal/mol)[1]			$\Delta\Delta G°_{H2O}$[2]				
Position	pH 5	pH 7	pH 9	pH 5	pH 7	pH 9	Slope[3]	pH_{mid}[5]
Δ+PHS	11.8	11.9	11.5	-	-	-	-	2.2
G20	7.5	8.9	8.7	-4.3	-3.0	-2.8	-0.7	3.1
V23	5.1	7.2	7.0	-6.7	-4.7	-4.5	-1.0	3.7
L25	4.0	5.5	5.3	-7.8	-6.4	-6.2	-0.7	3.8
F34	4.8	6.5	6.2	-7.0	-5.4	-5.3	-0.8	3.8
L36	5.2	7.1	6.9	-6.6	-4.8	-4.6	-0.9	3.8
L37	8.1	9.5	9.5	-3.7	-2.4	-2.0	-0.7	3.0
L38	10.1	10.7	10.6	-1.7	-1.2	-0.9	-0.3	2.4
V39	3.9	5.7	5.6	-7.9	-6.2	-5.9	-0.9	3.8
T41	8.7	9.2	9.1	-3.1	-2.7	-2.4	-0.2	2.9
A58	4.9	6.6	6.3	-6.9	-5.3	-5.2	-0.8	3.7
T62	7.8	9.5	-	-4.0	-2.4	-	-0.8	3.2
V66	6.9	7.8	-	-4.9	-4.1	-	-0.4	3.4
I72	6.5	8.9	-	-5.3	-3.0	-	-1.2	3.3
I74	4.1	6.1	-	-7.7	-5.8	-	-1.0	3.9
A90	3.5	5.3	-	-8.3	-6.6	-	-0.9	3.9
Y91	5.7	6.7	-	-6.1	-5.2	-	-0.5	3.6
I92	3.4	5.4	-	-8.4	-6.5	-	-1.0	4.0
V99	5.9	7.3	-	-5.9	-4.6	-	-0.7	3.6
N100	3.9	5.3	-	-7.9	-6.6	-	-0.7	4.0
L103	3.8	5.2	-	-8.0	-6.7	-	-0.7	4.0
V104	5.1	6.6	-	-6.7	-5.3	-	-0.7	3.6
A109	5.2	5.6	-	-6.6	-6.3	-	-0.2	3.5
N118	10.3	10.6	-	-1.5	-1.3	-	-0.1	2.4
L125	6.3	7.1	-	-5.5	-4.8	-	-0.4	3.6
A132	4.3[4]	4.6[4]	-	-7.5	-7.3	-	-0.1	-

[1] Data collected by chemical denaturation with GdnHCl monitored by Trp fluorescence at 25 °C in 100 mM KCl and 25 mM buffer. All $\Delta G°_{H2O}$ errors are 0.1.
[2] $\Delta\Delta G°_{H2O} = (\Delta G°_{H2O,His} - \Delta G°_{H2O,\Delta+PHS})$
[3] $(\Delta\Delta G_{pH7} - \Delta\Delta G_{pH5})/2$
[4] Data collected by monitoring CD signal
[5] Measured from acid unfolding monitored by Trp fluorescence

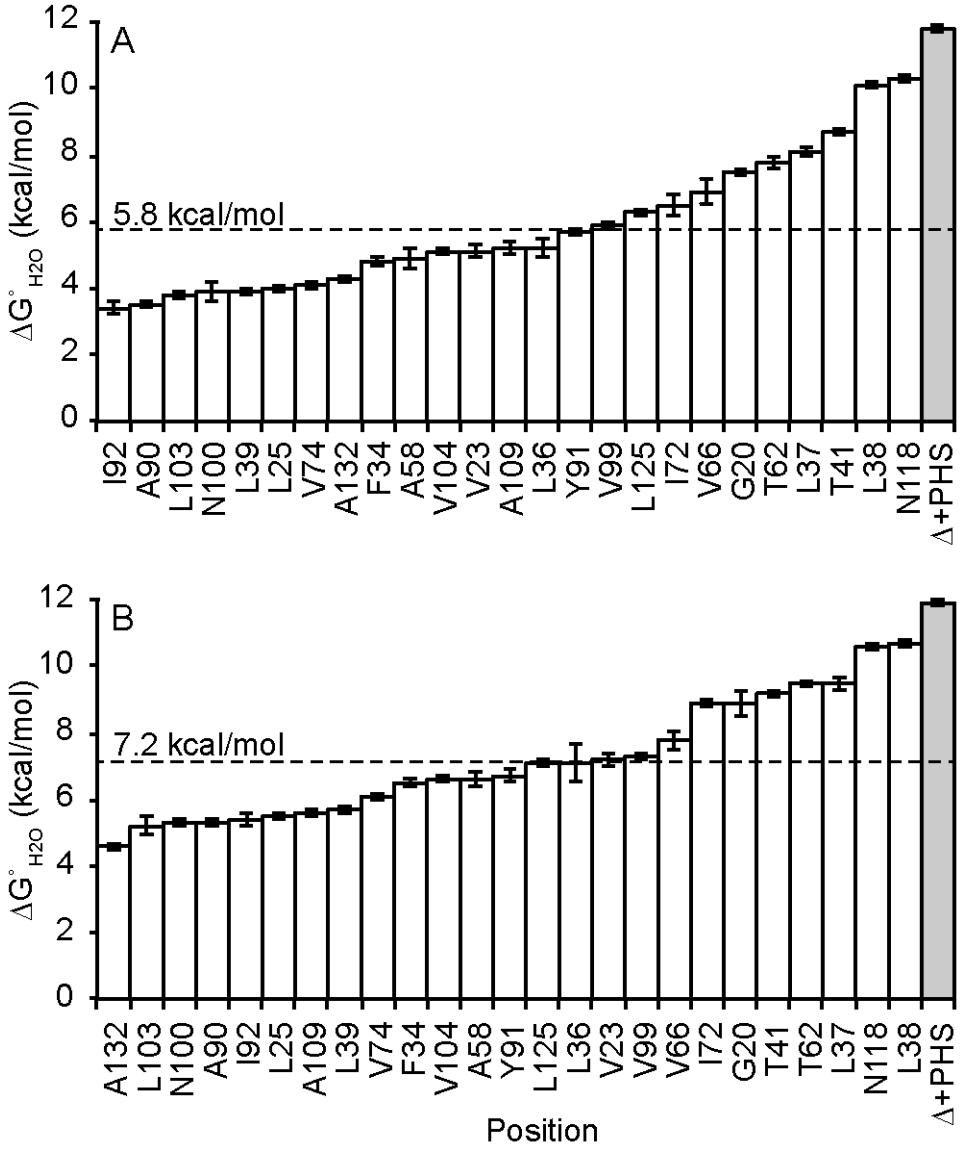

Figure 3.1 Thermodynamic stability of His variants at pH 5 and 7. Thermodynamic stability of His-containing variants (white) and Δ+PHS (grey) at pH 5 (top) and pH 7 (bottom). Error bars depict fit errors. The average stability of the His-containing variants is represented by a dashed line.

location in the protein. If the effects of stability are grouped in terms of structural element, what emerges is that the substitution with His is less destabilizing in loops than in β-strands or α-helices (Table 3.2).

3.4.2 Acid unfolding profiles

The acid unfolding monitored by Trp fluorescence for 24 of the 25 variants is shown in Figure 3.2. The midpoints of unfolding are included in Table 3.1. All the variants unfold at a pH higher than the parent protein, consistent with the decrease in thermodynamic stability at pH 7. NMR spectroscopy has been used extensively to measure pK_a values of histidine residues[103–105]. The reason NMR spectroscopy was not used to determine the pK_a values of the 25 internal His residues in SNase was that data on the pH dependence of stability described ahead suggested that the protein would unfold before the acid limit of the titration could be recorded, preempting the possibility of obtaining titration data from which pK_a values could be resolved accurately.

3.4.3 Ionization state of His residues

Linkage analysis of the pH dependence of $\Delta\Delta G°_{H2O}$ measurements can be used to determine the apparent pK_a of the buried His residues[1,49,60]. If the pK_a of the internal His were depressed relative to its normal pK_a in water (~6-6.5), then the $\Delta\Delta G°_{H2O}$ will decrease as the pH is lowered below the pH corresponding to the pK_a in water (Fig. 3.3).

The $\Delta\Delta G°_{H2O}$ will decrease at a rate of 1.36 kcal/mol for every unit shift in pK_a. Note that whereas the stability of the parent protein is insensitive to pH between pH 7 and

Table 3.2 Thermodynamic stability of His variants grouped by secondary structure. Average thermodynamic stability of SNase variants with His residues at internal positions organized according to the type of structural element where the internal His are found

Structural element	Average ΔG°_{H2O} [1]		Average $\Delta\Delta G^{\circ}_{H2O}$ [2]	
	pH 5	pH 7	pH 5	pH 7
α - helix	5.4 ± 1.4	6.7 ± 1.5	-6.3 ± 1.4	-5.2 ± 1.5
β - strand	4.7 ± 1.0	6.5 ± 1.1	-7.2 ± 1.0	-5.4 ± 1.1
other	7.7 ± 2.4	8.6 ± 2.1	-4.2 ± 2.4	-3.3 ± 2.1

[1] Errors given are standard deviations

[2] $\Delta\Delta G^{\circ}_{H2O} = (\Delta G^{\circ}_{H2O,\Delta+PHS} - \Delta G^{\circ}_{H2O,variant})$

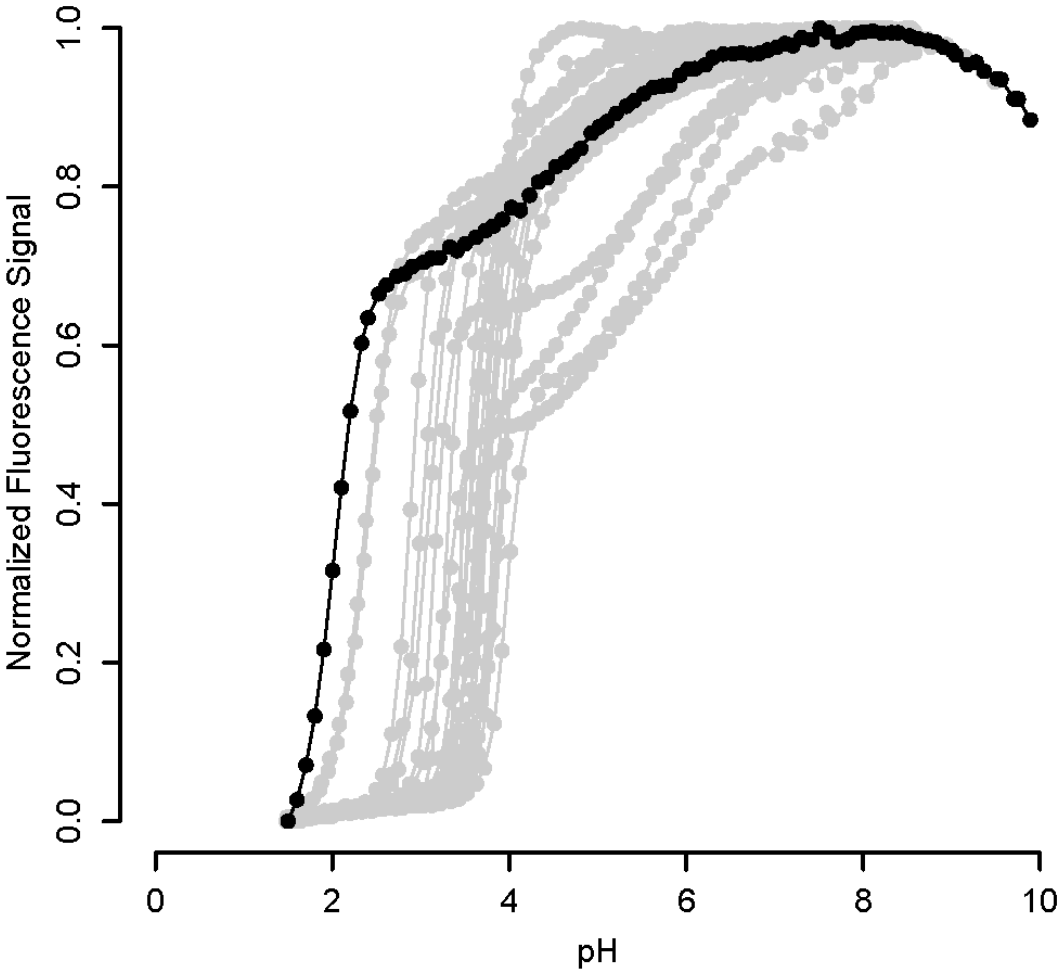

Figure 3.2 Acid unfolding of 25 His variants. Acid unfolding profiles monitored by Trp fluorescence of Δ+PHS (black) and His variants (grey).

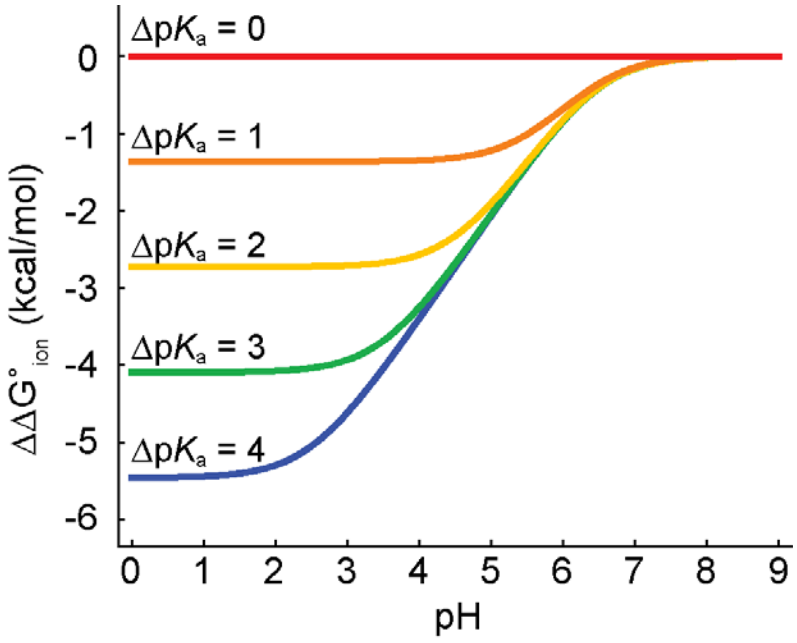

Figure 3.3 Simulations of the pH dependent component of thermodynamic stability of a His containing variant relative to the background protein ($\Delta\Delta G°_{H2O}$). These curves reflect how the stability would change as the pH decreases from an arbitrary reference $\Delta G°_{H2O}$ at pH 9. The different colors represent curves calculated with shifts in pK_a (ΔpK_a) of different magnitude: ΔpK_a = 0 (red), ΔpK_a = 1 (orange), ΔpK_a = 2 (yellow), ΔpK_a = 3 (green), and ΔpK_a = 4 (blue) units.

pH 5, the stability of all the His containing variants was lower at pH 5 than at pH 7. This is consistent with the buried His residues having depressed pK_a values.

At pH 5, the thermodynamic cost for a substituting His at an internal position ranged from 1.5 to 8.4 kcal/mol with an average $\Delta\Delta G°_{H2O}$ of -6.0 ± 2.0 kcal/mol. By comparing the value of $\Delta\Delta G°_{H2O}$ at pH 5 and at pH 7 it is possible to establish if the pK_a values of the His residues are normal or depressed. This required calculating the slope of $\Delta\Delta G°_{H2O}$ vs pH between pH 7 and pH 5 (Table 3.1). The slopes reported in Table 3.1 span from pH 5 to pH 7, encompassing the region of pH corresponding to the normal pK_a of His in water. As such, the reported slope will be shallower than the value of 1.36 kcal/mol/pH.

By comparing the experimentally calculated slopes between pH 7 and pH 5 to ideal slopes calculated from simulated ionization curves of a single His with different pK_a values in the native state (Fig. 3.3), it was possible to estimate the magnitude of the pK_a shift experienced in the 25 different single His substitutions. The slopes listed in Table 3.1 range from -0.1 to -1.2, which corresponds to an apparent pK_a shift of 0 to > 4 units. The data are fully consistent with all His residues having pK_a values lower than the normal pK_a in water. The pK_a values of His-41, His-118, His-125 and His-132 are depressed but only slightly. The majority of the other His residues appear to be experiencing substantial depression of their pK_a values. The positions with the most shifted pK_a values, such as 23, 72, and 92, are clustered in and around the β-barrel.

In conjunction with the experimentally determined pH_{mid} values of each variant (Table 3.1), it appears that in the majority of these variants the buried His residues remain neutral over a wide range of acidic pH and until the protein unfolds.

3.4.4 Far-UV CD

To examine how the presence of a His substituted at an internal position affects the global structure of the protein, far-UV CD spectra were collected for each variant at pH 7 and pH 5 (Fig. 3.4). For 24 of the 25 positions substituted with a His, the spectra were similar to that of the parent protein at both pH values. Only the variant with His-91 showed major differences relative to the parent protein and at both pH values. There is a relatively stronger signal at 208 nm than at 222 nm, which is opposite to the normal CD spectrum of SNase. This behavior is not unique to His substitutions as it has been seen with all other substitutions of Tyr-91. This is clearly a spectral artifact that is not indicative of structural reorganization[36,73].

3.4.5 NMR spectroscopy

NMR spectroscopy was used in an attempt to more directly determine the pK_a of the His side chain in the L36H, T62H, and V66H variants. ^{1}H-^{15}N HSQC spectra were measured over a range of pH from approximately pH 8 down to the pH where the protein begins to acid denature (Fig. 3.5). All three variants showed clear crosspeaks and chemical shift dispersion indicating that the proteins were folded, consistent with the CD and Trp fluorescence data collected at pH 5 and 7. At pH* ~7.5 and ~3.5, the collected spectra had high similarity to the reference protein.

To determine the ionization and tautomeric states of the introduced His, ^{1}H-^{15}N long range HSQC (LR-HSQC) spectra were collected over the same pH ranges for all three variants (Fig. 3.5). The LR-HSQC takes advantage of the long range (2- or 3-bond) J-couplings between the non-labile Hε1 and Hδ2 atoms and the Nδ1 and Nε2 atoms of

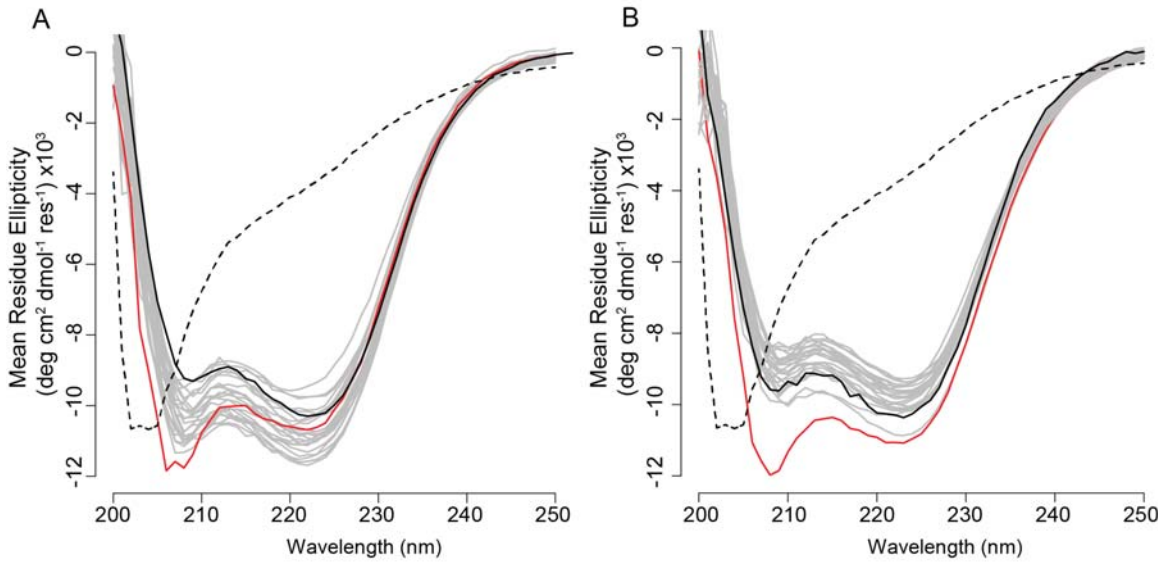

Figure 3.4 Far-UV CD spectra for variants with internal His residues. (A) Individual spectra for each His-containing variant at pH 5 are shown in gray. Spectrum for the variant with His is shown (red), the parent protein, Δ+PHS (black), and the constitutively unfolded T62P variant of wild type SNase (dashed) is shown for comparison. **(B)** Same as (A) but at pH 7.

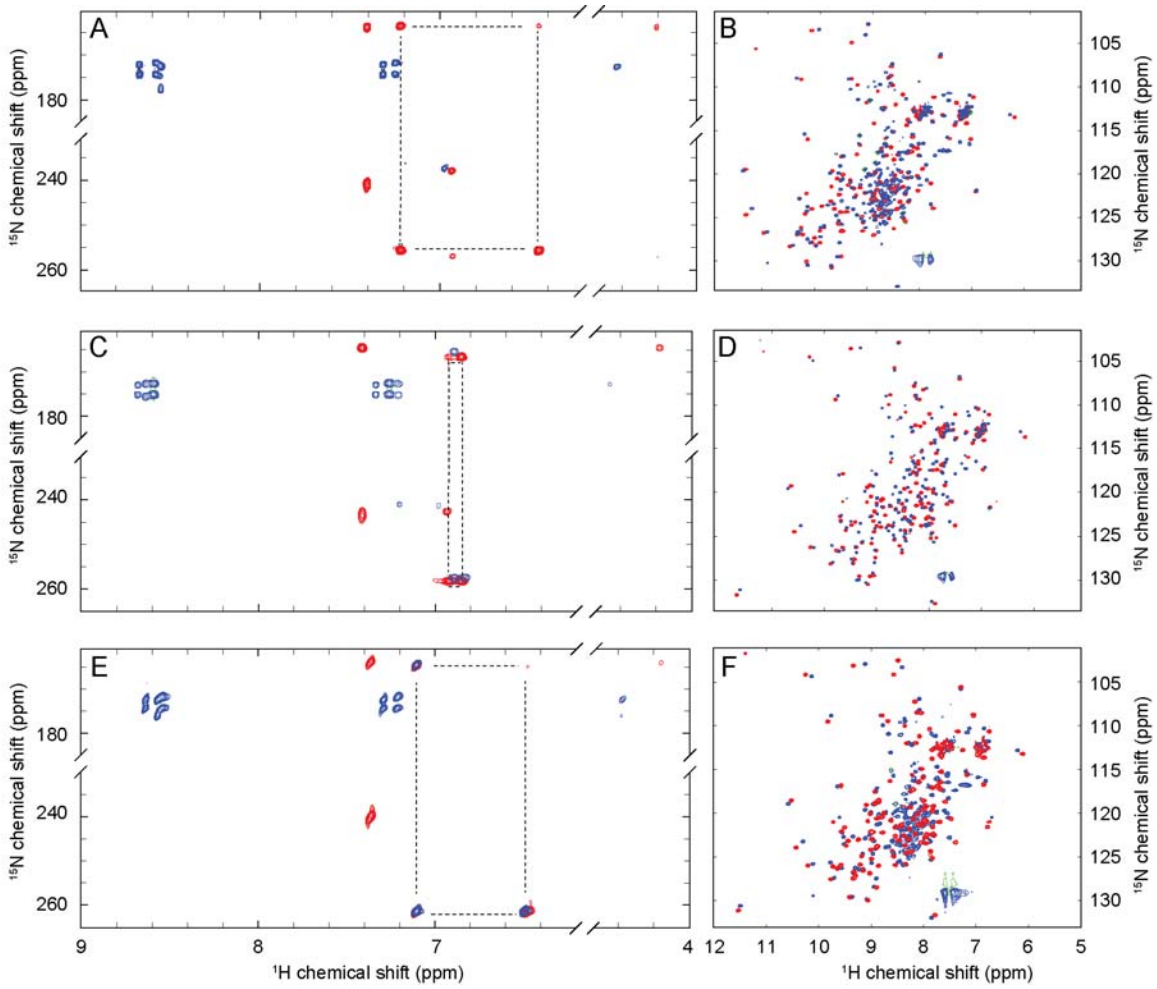

Figure 3.5 NMR spectra of His variants. HSQC (right) and LR-HSQC (left) experiments at high pH* (red) and low pH* (blue) for variants with His-36 (A,B), His-62 (C,D), and His-66 (E,F). HSCQ and LR-HSQC spectra were collected at low pH* 4.1, 3.7, 3.5 for His-36, His-62 and His-66, respectively. High pH* is ~7.5 for all variants.

the side chain imidazole[50]. When His is ionized, both Nδ1 and Nε2 are protonated (+>NH-type) and they resonate at ~174 and ~178 ppm, respectively[106]. In the neutral state, His may exist in one of two tautomeric forms, where only one of the two nitrogen atoms is protonated. (Fig. 3.6) In this state, the protonated (>NH-type) and deprotonated (>N:-type) nitrogens resonate at canonical values of ~168 and ~250 ppm, respectively, when the side chain exists in only one tautomeric state. The tautomeric states may be distinguished from one another by the characteristic "L" pattern, when Nδ1 is protonated, and upside-down "L" pattern, when Nε2 is protonated.

In all three variants examined, the His side chains show chemical shifts consistent with groups existing in the neutral state at all pH values examined. Furthermore, all three internal His residues adopt the same tautomeric state, with Nδ1 being protonated. These data are consistent with the equilibrium thermodynamic measurements that suggest internal His remains neutral until the protein globally unfolds.

3.4.5 Crystal structures

Crystal structures for eight of the 25 His variants were solved to determine what potential interactions in the microenvironment may be influencing the pK_a of the internal His residues. The structures for V23H, L25H, L36H, T62H, V66H, I72H, Y91H, and I92H variants were refined to resolutions 1.58 Å to 1.80 Å (Fig. 3.7, Table A2). All crystals were grown in buffer conditions in which the His residues are expected to be neutral. Overall, the tertiary structures of all eight variants did not deviate significantly from that of the parent protein (PDB ID: 3BDC), each with a C_α RMSD of 0.2 Å.

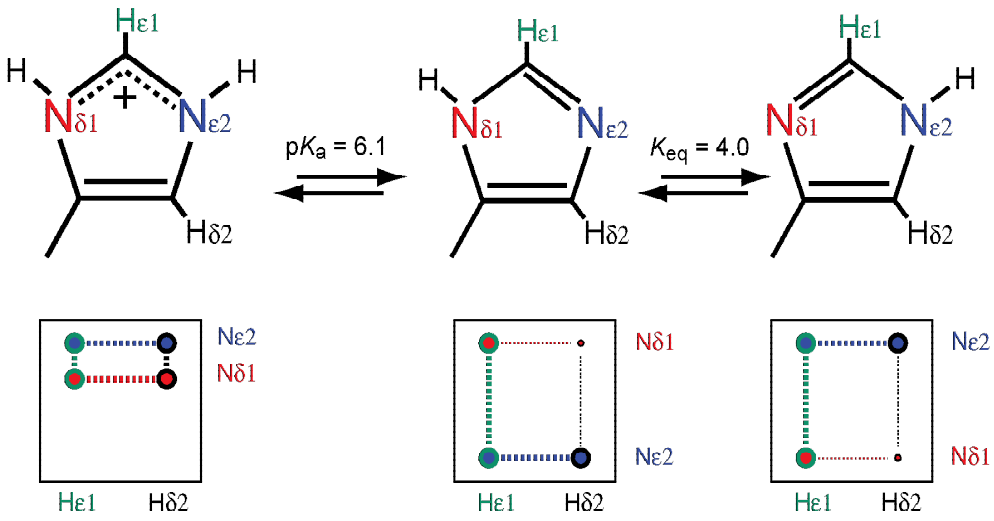

Figure 3.6 Histidine tautomers and associated LR-HSQC patterns. Diagram of all three imidazole forms of the His side chain (top) and the corresponding crosspeak patterns from long range HSQC experiments (bottom).

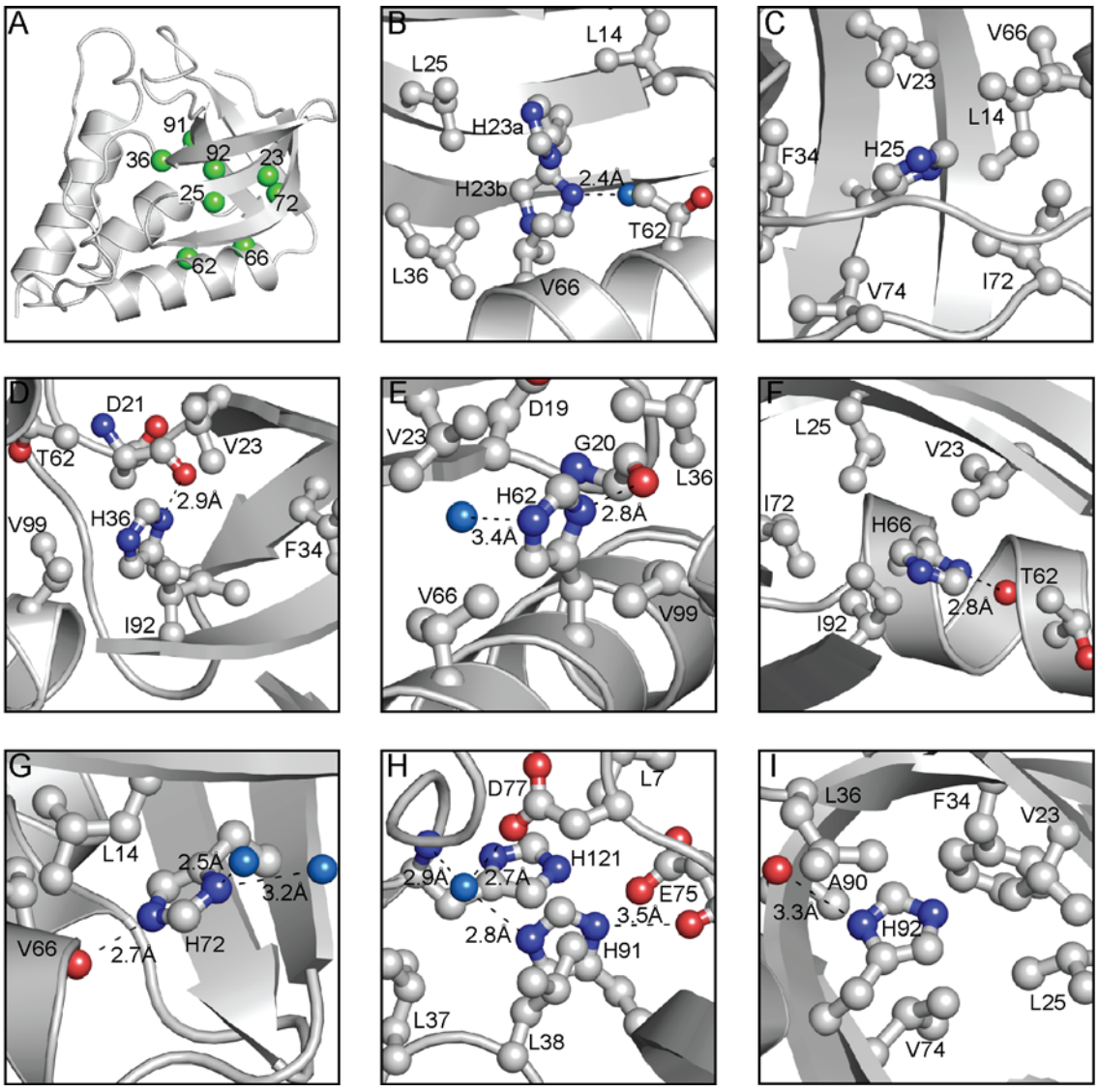

Figure 3.7 Microenvironments surrounding 8 buried His residues in SNase. (A) Crystal structure of Δ+PHS SNase showing the positions that were substituted with His. Close-up of the region surround His-23 (PDB ID 4ZUI) (B), His-25 (PDB ID 5C4Z) (C), His-36 (PDB ID 4LAA) (D), His-62 (PDB ID 5I9P) (E), His-66 (PDB ID 5C3W) (F), His-72 (PDB ID 4PNY) (G), His-91 (PDB ID 4ZUJ) (H) and His-92 (PDB ID 5C4H) (I). His side chain and nearby residues (ball and stick) and water molecules (blue spheres) are shown .

In the V23H variant the His-23 side chain is observed in two different conformations with equal occupancies, the A conformer pointing towards the naturally occurring microcavity in the β-barrel and the B conformer pointing towards α-helix 1. The A conformer has no potential hydrogen bonding partners within 4 Å of either the Nε1 or Nε2 atoms. The B conformer, however, has a water molecule situated 2.5 Å away from the Nδ1 atom (Fig. 3.7B). Using the WHAT IF software[107] to determine the solvent accessible surface area (SASA) of the side chain, the A and B conformations have 0.0% solvent accessibility.

The His-25 is completely buried in the middle of the β-barrel with 0.0% solvent accessible surface area (Fig. 3.7C). There are only 3 atoms within a 4 Å radius of either the Nδ1 or Nε2 atoms that could participate in a hydrogen bond with His-25. The backbone carbonyl oxygen atoms of Lys-24 and Thr-13 are 3.4 Å and 3.6 Å, respectively, away from the Nε2 atom and the backbone nitrogen of Leu-14 is 3.8 Å away from the Nδ1 atom of His-25 (Fig. 3.7C). However, given the geometries and distances between these atoms, there is no strong evidence for hydrogen bonding. Without evidence for a direct interaction, the side chain of His-25 appears to be dominated by the hydrophobic environment inside the β-barrel.

His-36 towards the microcavity in the β-barrel (Fig. 3.7D). Despite being completely inaccessible to bulk solvent, the Nδ1 of His-36 is within 2.9 Å of the carbonyl oxygen of Asp-21, which is normally unsatisfied. No other potential hydrogen bond donors or acceptors are observed near either nitrogen in the side chain of His-36.

The side chain of His-62 is internal and points into the microcavity in the β-barrel with 0.0% of the side chain accessible to water (Fig. 3.7E). A single water molecule is

visible 3.4 Å away from the Nε2 atom and 3.6 Å away from the Nδ1 atom. While being nearby, this water is situated above the plane of the imidazole which precludes the formation of a hydrogen bond but may be forming a hydrogen-π interaction. His-62 can form a hydrogen bond between the Nδ1 atom and the carbonyl oxygen of Gly-20, which is 2.8 Å away.

The side chain of His-66 is entirely buried within the β-barrel (Fig. 3.7F). There is only one potential hydrogen bonding partner, the carbonyl oxygen of Thr-62, which is 2.9 Å away from the Nδ1 atom of His-66. There were no visible distortions to α-helix 1 or the global secondary structure as were seen in previous spectroscopic data collected from Lys, Asp, or Glu-66 variants[108].

His-72 has the greatest solvent accessibility among the internal His residues that were studied. The side chain imidazole has an accessibility of 7.9%. It points towards the end of α-helix 1 (Fig. 3.7G). Both nitrogen atoms have potential bonding partners. The Nδ1 is 2.5 Å from a bulk water molecule and the Nε2 atom is 2.7 Å away from the carbonyl oxygen of Val-66. Given the distances and geometries of these atoms, they are most likely forming hydrogen bonds with the His side chain. Other ionizable side chains at positions 72 also have limited access to bulk water, which is intriguing because their thermodynamic properties are those of a completely buried ionizable group[52].

The side chain of His-91 points towards the loop region near the beginning of α-helix 3 and is inaccessible to solvent (Fig. 3.7H). In Δ+PHS, Tyr-91 is part of a hydrogen bond network where it acts as an donor to the side chain of Asp-77 and an acceptor to the backbone amide of His-121[50]. The side chain of His-91 overlays well with the orientation of Tyr-91. In the Y91H variant, the hydroxyl of Tyr-91 is replaced with a buried water

molecule that is 2.8 Å away from the Nε2 atom of the His-91. This water molecule forms the same hydrogen bonds with backbone of His-121 and the Oδ2 atom of Asp-77 as seen in Δ+PHS. The Nδ1 atom of His-91 is 3.2 Å and 3.5 Å, respectively, from the backbone carbonyl and Oε1 of Glu-75, although there is no evidence in the structure that His-91 forms hydrogen bonds with either moiety.

The side chain of His-92 is positioned in the β-barrel with 1.4% of the side chain solvent accessible. The Nδ1 atom of His-92 is 3.0 Å away from the backbone nitrogen of His-92 and 3.3 Å from the carbonyl oxygen of Ala-90 (Fig. 3.7I). However, given the geometries and that the carbonyl oxygen of Ala-90 is involved in a hydrogen bond with Leu-37, neither of these pairs would successfully sustain a hydrogen bond, leaving His-92 involved with no hydrogen bonds.

3.5 Discussion

3.5.1 Difficulties in measuring His pK_a values

It turned out to be unfeasible to measure the pK_a of the buried His side chains owing to the shift in the acid unfolding of the protein to high pH values that would preclude measurement of the full titration curve. On the other hand, the pH dependence of stability is consistent with 19 of the 25 internal His residues having substantially depressed pK_a values with shifts considerably larger than the shifts usually observed with surface His residues.

Given the errors intrinsic to the $\Delta G°_{H2O}$ values measure and the limited range of pH where stability was measured, it was impossible to determine the pK_a value of the His residues in the manner that has been done for Lys, Asp, and Glu[2,36,61]. However, the data

are sufficient to provide boundary value. Using basic linkage analysis (Eq. 3.1) to simulate the $\Delta\Delta G°_{H2O}$ of the protein as the pH changes (Fig. 3.3), it is possible to estimate the range in pH where the pK_a of the buried His falls.

$$\Delta\Delta G^{\circ}_{H2O} = \Delta\Delta G_{mut} - RT \ln\left(\frac{1+e^{2.303n(pK_a^D - pH)}}{1+e^{2.303n(pK_a^N - pH)}}\right) \qquad \text{(Eq. 3.1)}$$

Using the simulations of the type shown in Fig. 3.3, it was possible to test different pK_a^D values (pK_a in the unfolded state) and determine the pK_a^N corresponding to the pK_a of the His residue in the native state. For position 132, which has a slope of -0.1, regardless of the value used for pK_a^D (ranging from 6.0 to 6.5) the shift in the pK_a^N was -0.2, which is consistent with His-132 titrating with a normal pK_a. Similarly His-125 has a slope of -0.4 that corresponds to a maximal shift in pK_a^N of -0.5, which is measurable but small.

The cases of His-62 and His-66 are more instructive as considerable information was derived from the NMR spectroscopy studies. NMR studies indicated that both His-62 and His-66, which in the thermodynamic studies have slopes of -0.8 and -0.4, respectively (Table 3.1), remained in the neutral state down to pH 3 where acid unfolding began (Fig. 3.5). The simulation for His-66 predicts a pK_a^N of < 4.8 regardless of the input for the pK_a^D value. When using a value of 6.5 for the pK_a^D variable for His-62, the resulting pK_a^N is < 5 while setting the pK_a^D value to 6.0 yields a pK_a^N of < 4. This demonstrates that the analysis of the slopes of stability with pH can be interpreted in a reliable manner to contribute boundary values to pK_a of the His residues. A wealth of data on the internal Lys, Glu, and Asp residues at these 25 internal positions further

corroborate that validity, accuracy, and precision of the determination of pK_a values through this thermodynamic approach.

3.5.2 Structure correlation with pK_a shifts

To identify the factors that might be influencing the pK_a of the buried histidines, the $\Delta\Delta G°_{H2O}$ values and pH sensitivity of the 25 variants were analyzed in terms of the conformation of the backbone to which the internal His residues are connected. Each amino acid can be found in elements of different secondary structure (α-helix, β-strand, coil) with characteristic frequency. The structural preference of histidine is somewhat complex. In three independent surveys, histidine has been shown to be weakly correlated with all secondary structural elements[74], positively correlated with helices and negatively correlated with β-strands[109], or favorably correlated with both α-helices and β-strands[110].

For SNase, internal positions 20, 37-41, and 109-118 can be classified as coil positions, positions 58-66, 99-104, and 125-132 are α-helix positions and 23-26 and 72-92 are β-strand positions. Using these groupings, the average $\Delta\Delta G°_{H2O}$ values measured at pH 7 were -3.3, -5.2 and -5.4 kcal/mol, respectively. Similarly, the average $\Delta\Delta G°_{H2O}$ values at pH 5 were -4.1, -6.4 and -7.1 for coil, helix, and strand respectively (Table 3.2). This agrees best with the findings from Levitt[110]. At both pH values, the His substitution is most destabilizing in the strand and helix positions. The average $\Delta\Delta G°_{H2O}$ values in the coil regions for the substitutions to His at pH 7, where the side chain is assumed to be neutral, is similar as reported for neutral ionizable and other polar substitutions at the same positions. This suggests that His is not unique in the decrease in thermodynamic stability when neutral.

The same analysis was done to determine if the same correlations carry over from $\Delta\Delta G^{\circ}_{H2O}$ to the shift in the pK_a as determined by the calculated slope. The average slope for His residues in coil regions was -0.4 ± 0.3, -0.6 ± 0.2 for helix positions, and -0.9 ± 0.2 for positions in β-strands. The same trend is observed indicating that the less-flexible α-helix and β-strand elements of secondary structure lead to a larger shift in the pK_a value of the His, consistent with the previous observation that although it is more destabilizing to substitute core positions in the β-barrel with ionizable groups, the residues buried in the barrel exhibit the largest shifts in pK_a values[2,61].

3.5.3 Microenvironments of buried His side chains

The Lys and Arg residues that have been studied at these internal positions are very different from His in the length, chemistry and conformational flexibility of their side chains. His is inherently more constrained conformationally. In all eight crystal structures that were examined the His side chain is buried; in five cases the side chain is completely inaccessible to solvent. This is not entirely surprising. Unlike the Lys and Arg side chains, buried His side chains, even completely buried ones, are not unusual[97]. What the present, systematic study with His at 25 internal positions does demonstrate is that burial of the bulky His side chain is relatively easy to achieve, neither requiring special structural adaptations nor inducing any significant structural reorganization. Although the His side chains appear in most of the structures to interact with potential hydrogen bonding partners, the thermodynamic data suggest that these hydrogen bonding interactions are not sufficient to offset the consequence of the dehydration experienced upon burial.

Some authors have speculated about a possible correlation between solvent accessibility, burial in the hydrophobic interior of the protein, and pK_a values[93,97,111]. Edgcomb and Murphy surveyed 37 histidines from 13 proteins and concluded that there is no correlation between the side chain burial and depressed pK_a values. A computational approach to predicting pK_a values of histidines also arrived at a similar conclusion: there is no correlation between ASA and pK_a values and thus accessible surface area can be ignored with no ill effects[93]. For the set of His variants presented here, the untitratable histidines at positions 36, 62 and 66 all are completely buried. These findings do not directly contradict earlier studies, however, as other untitratable histidines were found to be among the most buried[26,97].

In a large survey of hydrogen bonds involving all amino acids, it was found that His made 0.8 hydrogen bonds on average as a donor and 0.3 as an acceptor suggesting that His need only satisfy one of its possible side chain hydrogen bond interactions[90]. A more recent survey of histidines observed that all His side chains with < 30% ASA made at least 1 interaction and side chains that made no interactions were predominantly > 60% ASA[97]. This is contrary to what is observed with the 8 crystal structures with buried His side chains with < 10% accessible surface area capable of forming 2, 1, or even no hydrogen bonding interactions. It is important to note that the His residues we have studied were introduced mutagenically while the previous studies examined natural His residues. It is possible that the properties of naturally occurring His residues reflect evolutionary adaptation. The structures of the internal His residues in SNase demonstrate that His side chains can exist in an extremely hydrophobic environment without the need

of being involved in any polar contact. The same has been observed most dramatically for buried Lys side chains[2] but also for buried Asp and Glu side chains[60,61].

3.5.4 Comparison between internal His, Lys, Arg, Asp, and Glu

Remarkably, the average decrease in thermodynamic stability of SNase upon substitution of an internal position with any ionizable residue is comparable when measured at a pH near the normal pK_a of the residue. Substitution of neutral His appears to be comparable to previously studied ionizable groups buried at the same 25 internal positions in the neutral state. At pH 10, the average cost for substituting neutral Lys is -4.8 ± 2.4 kcal/mol; at pH 5 the average cost for substituting neutral Glu and Asp is -4.4 ± 1.5 kcal/mol and -5.1 ± 2.1 kcal/mol, respectively[2,60]. A similar comparison to Arg was not possible as the Arg side chain is never observed in the neutral state in the pH range where SNase is folded[1].

For substitution with His at pH 7 the average loss of stability is 4.7 ± 1.8 kcal/mol which is the same as the loss of stability upon substitution with Asn or Gln[36]. The striking similarity in the average loss of stability upon substituting internal positions with a polar residue suggests that the effect is dominated by differences in hydration of the polar side chain in water and in the protein. That is the only factor that all side chains have in common. The other factors that could be reflected in these numbers, such as differences in the extent of van der Waals interactions and conformational entropy of the side chain, are probably minor relative to the contributions from hydration. It should be remembered that these averages are averages over a very broad distribution. This suggests that the

polar side chains are sensitive to the characteristics of the unique microenvironment surrounding the side chain.

The more notable difference between the buried ionizable residues is between the properties of Arg and the properties of Lys, Asp, Glu, and His. We have been unable to measure shifts in pK_a values of buried Arg residues. This probably reflects a combination of factors. First, the normal pK_a of Arg in water is closer to 14 than to the value of 12 that is assumed by many[40]. Second, the charge in the guanidinium group is highly delocalized over 4 atoms and 3 bond, therefore the hydration energy of the charged guanidinium moiety is likely low thus the penalty for removing the group from water and sequestering it in a dry environment is also quite low. The Arg side chain is long and this helps it compensate for loss of hydrogen bonds with water by placing the buried guanidinium moiety where it can satisfy its 5 hydrogen bonds[1]. In contrast, the hydration energy of the amino group of the Lys side chain is high causing the transfer of the side chain from water into the relatively hydrophobic environment in the protein interior to be unfavorable. This is reflected in the substantial shifts of up to 5 pK_a units, always in the direction the promotes the neutral form of Lys[2]. Glu and Asp behave somewhat like Lys, and although the charge in the carboxylic group is more delocalized than in the amine in Lys, the carboxyl group is well hydrated and transfer to a hydrophobic environment is unfavorable.

Despite the fact that the imidazole side chain of the His residue is aromatic, in principle more compatible with hydrophobic environments, and that the charge can also be delocalized in the ring, the ionization properties of His residues buried in hydrophobic environments are more similar to Lys than to Arg. Although the exact pK_a values of the

internal His residues could not be determined with accuracy or precision in this study, the data show conclusively that they are shifted, in most cases depressed by more than 1 or 2 pK_a units. This is a conservative estimate. In the case of His-36, His-62 and His-66 it was possible to show conclusively with NMR spectroscopy that the pK_a values are depressed by at least 3 units; the side chains of these His residues remain in the neutral state under highly acidic conditions and only become protonated concomitant with the acid unfolding transition.

The structures of variants with internal Lys, Arg, and His residues in SNase reveal some potentially interesting differences in the capacity of these residues to establish hydrogen bonds. Even when buried in the deepest and most hydrophobic regions of the core of SNase, the Arg side chain has the ability to reach towards the protein surface and find backbone atoms or enhance water penetration to satisfy its multiple hydrogen bonds[1]. The Lys side chain can often be found buried in totally hydrophobic environments[52]. The conformation of the side chain observed in crystal structures does not seem to be influenced or driven by hydrogen bonding interactions. The His side chain is somewhat like Lys. It appears to form hydrogen bonds more readily than Lys, but it can be found buried without its hydrogen bonds fully satisfied.

There is no correlation between the effects reported by internal acidic or basic residues buried at the same position in SNase. Even the effects of different acidic or basic residue on the stability of the protein are not comparable. It is very clear that the nuances of interactions of the different side chains with their microenvironments determine many of their properties. The properties of buried His residues appear to be especially sensitive to ion pairing and hydrogen bonding interactions.

Chapter 4

Anomalous pK_a values of buried ionizable residues in proteins can be sensitive to global stability

4.1 Abstract

Ionizable groups buried in hydrophobic environments in proteins are essential for all forms of biological energy transduction. It is increasingly apparent that conformational reorganization coupled to the ionization of the buried group is a major determinant of the pK_a values of these internal groups. In the presence of a buried group with an anomalous pK_a, the energy gap between the fully folded and local or partially unfolded states decreases as the pH of solution approaches the pH where the internal group becomes ionized. The ionization event is thermodynamically coupled to a shift from the fully folded state to local or partially unfolded states in which the nascent charge can become hydrated or gain access to an environment where the charge can be solvated by the protein. These alternative conformational states, not normally populated, can become the new ground state under conditions of pH conditions where an internal ionizable group is charged. If the ionization of an internal group promotes the transition to a new conformational state then its pK_a should be sensitive to the global thermodynamic stability ($\Delta G°_{H2O}$) of the protein because that determines the energy gap between the fully folded and all other states of the protein. This concept was tested by measuring the pK_a of two internal Lys residues in variants of staphylococcal nuclease with thermodynamic stabilities ranging from 4.9 to 13.8 kcal/mol. The pK_a of the internal Lys residues could be shifted by almost 2 pK_a units depending on the $\Delta G°_{H2O}$ of the parent protein. This confirms that the pK_a values of these Lys residues are determined by the probability of structural reorganization more than by local dielectric properties of the protein microenvironments of the ionizable moieties. These observations imply that structure-based pK_a calculations for buried groups and other electrostatic processes in

hydrophobic environments require accurate treatment of conformational reorganization, which remains an extremely challenging proposition.

4.2 Introduction

Ionizable residues are essential for many biological functions. Most ionizable residues are at the surface of proteins, but a few biologically important ones are buried in the hydrophobic interior of proteins. These internal ionizable residues play significant roles in processes such as H^+ transport[18,19], ion homeostasis[62], and catalysis[7,8,112]. To understand how these residues determine biological function it is imperative to know their pK_a values and to understand their molecular determinants.

The protein interior is much less polar and polarizable than water. Therefore, when an ionizable residue is sequestered from bulk water in the protein interior, the pK_a of the ionizable residue shifts relative to the normal pK_a in water, sometimes by several pK_a units, in the direction that promotes the neutral state[2,35,60]. The ionization of an internal residue often triggers structural reorganization to varying degrees owing to the change in the relative stabilities of alternate conformations[53,72,108,113,114]. The measured or apparent pK_a is actually a population-weighted average of the pK_a in the many different conformational states that exist in solution; therefore, these conformational changes influence the apparent pK_a value of the internal ionizable residue[53,115,116]. Conformational heterogeneity is important for protein function as it is often the alternative, accessible conformational states that are not the lowest energy state may still be the biologically active ones[117–119]. Understanding how conformational states other than the fully folded state affect the measured pK_a of an internal ionizable group is of inherent interest. The

problem is that the alternative conformational states are difficult to observe with direct structural methods[120].

Structure-based pK_a calculations have contributed important insights into possible molecular determinants of pK_a values, but accurate calculation of pK_a values is still extremely challenging, even for surface ionizable residues[34,52,114]. Calculation of the anomalous pK_a value for internal ionizable residues is even more challenging[121] because several factors that have been observed experimentally to affect pK_a values of internal groups are difficult to reproduce computationally: strength of interactions between the ionizable group[122], between the ionizable group and other polar moieties[122], interactions of buried ionizable residues with surface charges[54], and fluctuations between alternative conformational states[52,53]. The problem related to the roles of alternative conformational states as determinants of pK_a values is especially challenging in structure-based calculations. To reproduce this correctly, improvements must be made to sampling methodology the correctly predict alternative states and to force fields so the relatively small free energy differences between conformational states can be calculated accurately.

The direct detection of alternative conformational states of proteins with structural methods is challenging because they are transiently or marginally populated. In the case of alternative states promoted by the ionization of internal residues, this is less of a problem because exploring various conditions of pH makes it possible to find conditions where an alternative state becomes the new ground state. Here we seek to demonstrate the role of alternative conformations of proteins as determinants of the anomalous pK_a values of buried groups by showing that the apparent pK_a of an internal Lys residue can be shifted by secondary mutations far from the ionizable group that affect the global

thermodynamic stability ($\Delta G°_{H2O}$) of the protein.

$\Delta G°_{H2O}$ is the free energy gap between the folded and unfolded states. This is highly pH sensitive in proteins with buried ionizable groups with anomalous pK_a values. As the energy gap decreases the probability of populating alternative states increases. A decrease in the energy gap should affect the apparent pK_a of the internal ionizable residue because it will increase the population of alternative states in which, presumably, the anomalous pK_a in the fully folded form becomes more normal because it can contact water. The hypothesis we have examined is that the anomalous pK_a of a Lys in a protein will shift towards a more normal value in response to mutations far from the Lys that lower $\Delta G°_{H2O}$ of the proteins (Fig 4.1).

To test this hypothesis, Val-23 and Leu-36 in staphylococcal nuclease (SNase) were substituted with Lys. Previous experiments with a highly stable form of SNase showed that Lys-23 and Lys-36 titrate with pK_a values of 7.3 and 7.2, respectively. Both the V23K and L36K variants are also moderately stable at pH 7, with $\Delta G°_{H2O}$ = 4.6 and 4.7 kcal/mol, respectively[2]. Crystal structures of these two variants have been solved under conditions of pH in which the Lys residues are neutral and under these conditions the Lys side chains are completely buried and buried in the same general region of the protein[52]. The substitutions to Lys have now been done in five variants of SNase with different global stability (Fig. 4.2, Table 4.1). These stable variants are almost totally insensitive to pH in the pH range 5 to 10 and their thermodynamic stabilities bracket those of the original protein used to measure the pK_a values of 7.3 and 7.2. The pK_a values of Lys residues in all these different background were measured through linkage analysis of the pH dependence of stability (Fig. 4.3) as shown previously[49,100].

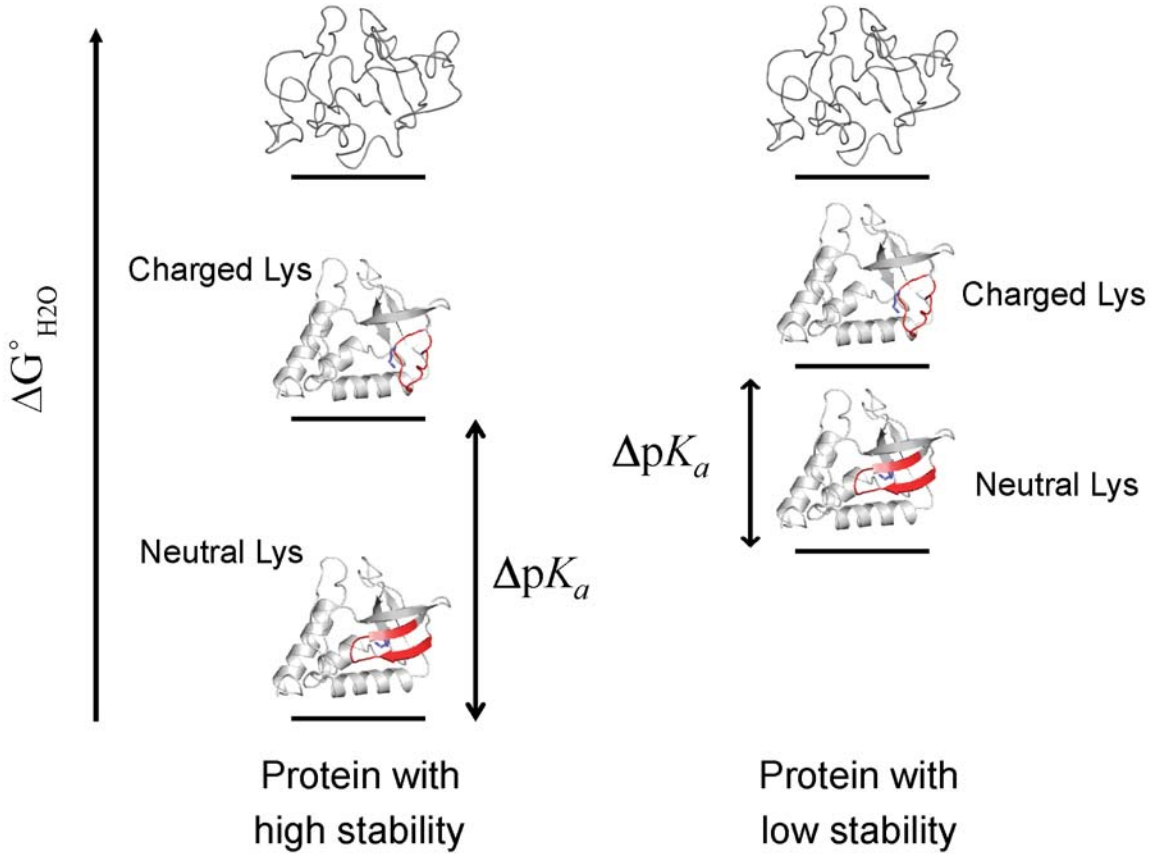

Figure 4.1 Relationship between energy gaps of protein states and apparent pK_a. Energy diagram showing the relationship between the magnitude of the energy gap between folded, partially unfolded and fully unfolded SNase and the magnitude of the apparent pK_a.

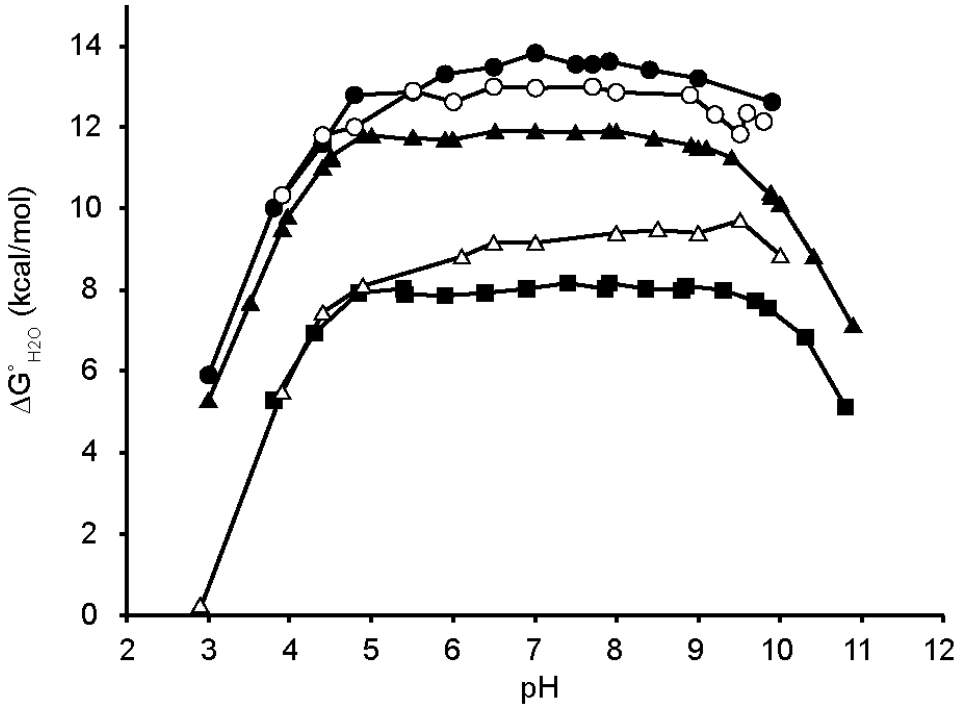

Figure 4.2 pH dependence of stability of five background variants of SNase. Thermodynamic stability of GLA (■), NVIAGA (△), Δ+GLA (▲), Δ+VIAGLA (○), and Δ+NVIAGLA (●) measured at various pH values. All measurements were taken at 25℃ and 100 mM salt.

Table 4.1 Thermodynamic stability of wild type SNase and of 5 stable variants[1]

Protein	$\Delta G°_{H2O}$[2]	Substitutions
WT	4.9	-
GLA[3]	8.0[4]	P117G, H124L, S128A
NVIAGA	9.2	D21N, T33V, T41I, S59A, P117G, S128A
Δ+GLA[3]	11.9[5]	del 44-49, G50F, V51N, P117G, H124L, S128A
Δ+VIAGLA	13.0	del 44-49, G50F, V51N, T33V, T41I, S59A, P117G, H124L, S128A
Δ+NVIAGLA	13.8	del 44-49, G50F, V51N, D21N, T33V, T41I, S59A, P117G, H124L, S128A

[1] Measured with GdnHCl titration monitored with Trp fluorescence at pH 7, 25℃ in 100 mM KCl and 25 mM HEPES
[2] $\Delta G°_{H2O}$ in kcal/mol
[3] Also known as PHS or Δ+PHS
[4] Value from Dwyer et al. 2000
[5] Value from Isom et al. 2010

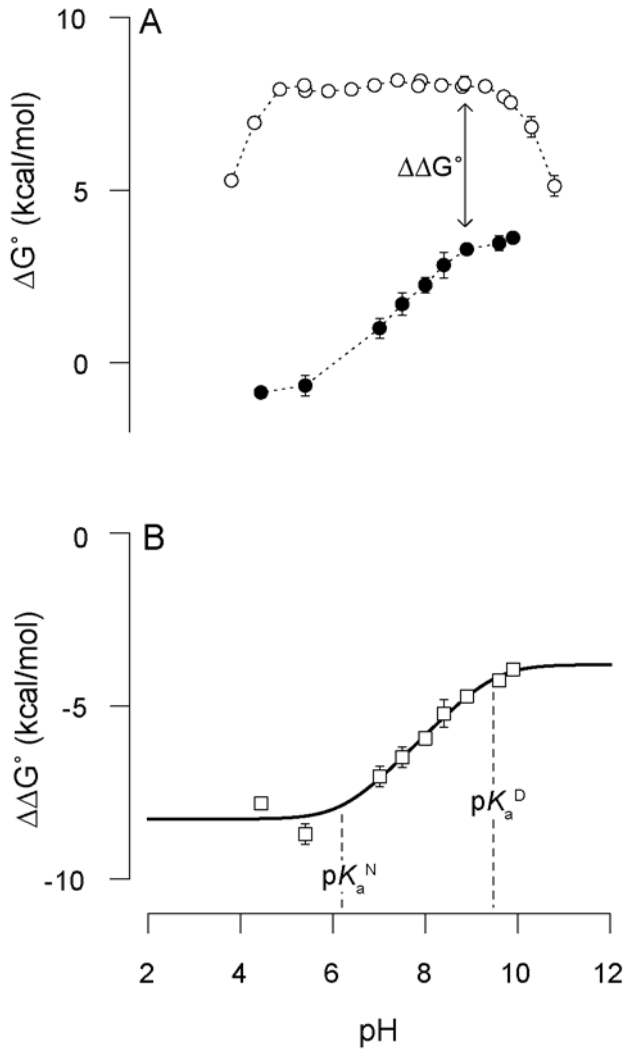

Figure 4.3 Linkage analysis of the pH dependence of stability for measurement of pK_a values. (A) Thermodynamic stability ($\Delta G°_{H2O}$) of background protein, GLA, (○) and Lys-containing variant, GLA/L36K, (●) with dotted lines to guide the eye. **(B)** Difference in thermodynamic stability ($\Delta\Delta G°_{H2O}$) between background and Lys-containing variant (open squares). Solid line represents fit with Equation 1.

The results of this study are fully consistent with the idea that the anomalous pK_a values of internal ionizable residues are affected by the global stability of the protein and that alternative conformations are important determinants of these pK_a values. The relationship between stability and pK_a turned out to be more complex than anticipated. These results show that accurate prediction of alternative conformations of proteins and the free energy difference between conformational states will be necessary for accurate prediction of pK_a values of buried ionizable groups. Efforts are on-going to describe the nature of conformational transitions that are coupled to the ionization of internal residues in detailed structural terms.

4.3 Materials and methods

4.3.1 Proteins

The Stratagene Quickchange kit was used to engineer the variants. Each protein was expressed in *E. coli* BL21/DE3 cells (Invitrogen) transformed with the plasmid Pet24a+. Proteins were expressed and purified by the method of Shortle and Meeker[69] as modified by Byrne et al.[70]

4.3.2 Equilibrium thermodynamic measurements

The Gibbs free energy of unfolding ($\Delta G°_{H2O}$) was measured using the intrinsic fluorescence of Trp-140 to monitor unfolding, as described previously[71]. The far-UV signal at 222 nm was used for His-132 variant. GdmCl (UltraPure grade, Invitrogen Life Technologies) was used as a denaturant, as described previously[71]. All measurements were performed with an ATF-105 automated fluorometer (Aviv Inc.) at 25℃. The protein

concentration in these experiments was 50 µg/mL in a buffer consisting of 100 mM NaCl with either 25 mM KAc, MES, HEPES, TAPS, CHES, or CAPS as appropriate for the pH.

4.3.3 Linkage analysis for determination of pK_a values

The pK_a values of ionizable groups are sensitive to their environment, thus for buried residues the pK_a values are very different in the folded and unfolded forms of the protein. When the protein is unfolded the ionizable groups are extensively hydrated and titrate with the normal pK_a values of model compounds in water. In the folded state the pK_a values of the folded groups are dependent on the location of the ionizable moiety and on nuances in the microenvironment such as the polarity and polarizability of the surrounding atoms. The pK_a values of internal groups are shifted relative to the pK_a values in water, always in the direction that favors the neutral state because the protein interior is not as good a solvent as water. Buried ionizable groups that titrate with anomalous pK_a values have a striking effect on the pH dependence of thermodynamic stability ($\Delta G°_{H2O}$). The thermodynamic contribution associated with the ionization of a single ionizable group ($\Delta G°_{ion}$) is described in Eq. 4.1

$$\Delta G^{\circ}_{ion} = -RT \ln\left(\frac{1+e^{2.303z\,(pH-pKa^{D})}}{1+e^{2.303z\,(pH-pKa^{N})}}\right) \qquad \text{Eq. (4.1)}$$

where z is the charge of the ionizable side chain, pK_a^D is the pK_a value of the group in the unfolded state and pK_a^N is the pK_a value in the folded state. For the ionizable group to have an impact on the stability of the protein, the pK_a^D and pK_a^N values must be different.

The background proteins used in this study display negligible pH dependence in $\Delta G°_{H2O}$ over the pH range from 5 to 10. A substitution to Lys at an internal positions has a marked destabilizing effect on the protein, even under conditions of pH where the Lys side chains is normally neutral. On account of the depressed pK_a the $\Delta G°_{H2O}$. becomes highly pH sensitive. The pH dependence of the difference in stability between the background protein and the variant with the internal Lys with anomalous pK_a can be analyzed with this function to obtain the pK_a of the internal Lys in the native state pK_a^N.

$$\Delta\Delta G^\circ = \Delta\Delta G^\circ_{mt} - RT \ln\left(\frac{1+e^{2.303z\,(pH-pKa^D)}}{1+e^{2.303z\,(pH-pKa^N)}}\right) \quad \text{Eq. (4.2)}$$

where $\Delta\Delta G°_{mt}$ reflects the pH-independent change in free energy for substituting Lys when the side chain is neutral. The thermodynamic linkage approach to monitoring pK_a values has been validated extensively with other equilibrium thermodynamic measurements[49,100,123,124] as well as NMR spectroscopy[55,125].

In cases where there is more than one ionizable group with an abnormal pK_a, the pK_a values can be extracted by fitting with Eq. 4.3, which is an extension of Eq. 4.2:

$$\Delta\Delta G^\circ = \Delta\Delta G^\circ_{mt,1} + \Delta\Delta G^\circ_{mt,2} - RT \ln\left(\frac{1+e^{2.303z\,(pH-pKa_1^D)}}{1+e^{2.303z\,(pH-pKa_1^N)}}\right) - RT \ln\left(\frac{1+e^{2.303z\,(pH-pKa_2^D)}}{1+e^{2.303z\,(pH-pKa_2^N)}}\right)$$

Eq. (4.3)

4.3.4 Far-UV CD Spectroscopy

An Aviv model CD-420 circular dichroism spectrophotometer was used to collect far-UV CD spectra. Samples consisted of 50 μg/mL protein, 100 mM KCl and either 25 mM TrisHCl. (pH 7) or 25 mM CAPS (pH 10). Spectra were collected for a protein sample in a 0.2 cm path-length quartz cuvette at 25 ℃. Measurements were taken at wavelength intervals of 1 nm using an averaging time of 3 s.

4.3.5 X-ray Crystallography

Δ+VIAGLA/V23K was crystallized by the hanging drop vapor diffusion method at 4 ℃. The reservoir solution consisted of 25% (v/v) 2-methyl-2,4-pentanediol in 25 mM potassium phosphate buffer (pH 8). Two molar equivalents of the inhibitor of pdTp and 3 molar equivalents of $CaCl_2$ were added to the protein solution before it was mixed with an equal volume of reservoir solution. The pdTp was synthesized in our laboratory[123].

Diffraction data were collected at 100K using a Bruker Duo Apex diffractometer. Initial phasing for all structures was obtained by maximum likelihood-based molecular replacement method with PHASER software within the CCP4 suite using a previously solved structure for Δ+PHS (PDB ID: 3BDC) as a search model. Prior to molecular replacement, 3BDC.pdb was modified by truncating residues 8, 23, 31, 41, 50, 51, 59, 80, 113-116, 123, 131 to Ala, removing residues 16-24, removing all water molecules, and resetting all B-factors to 20.0 $Å^2$. Model building using Coot and refinement with Refmac5 were performed iteratively to yield the final models. R_w and R_f residuals were monitored throughout the refinement. Water molecules were added during model

building to reflect spherical electron density in $2F_o$-F_c maps that were within 3.5 Å of a hydrogen bonding partner in the protein model. Final checks of the structures were done using SFCHECK and PROCHECK programs.

4.4 Results

4.4.1 Thermodynamic stability

This study was based on five variants of SNase that were engineered to be more stable than the wild type protein (Table 4.1). The naming convention used here is detailed in Table 4.1 for each variant. The "Δ" refers to the deletion of the Ω-loop (residues 44-49) and the "+" refers to the pair of mutations at positions 50 and 51. All other letters are taken from the residue that has been substituted into the protein (e.g. G = P117G). The thermodynamic stability measured with chemical denaturation showed that the most stable variant is almost three times as stable as the wild type (Table 4.1). This study uses five different background variants of SNase created by the incorporation of different combinations of stabilizing mutations[126,127]. On-going work in this laboratory has shown that the backbone structure of these proteins is superimposable.

Positions 23 and 36 were identified as useful starting points for this study based on previous experiments in the Δ+GLA background and crystal structures showing that the Lys amino group of the two variants are in similar environments[52]. The pK_a values for Lys-23 and Lys-36 measured previously are 7.3 and 7.2, respectively[2]. As the stability of the Δ+GLA/V23K and Δ+GLA/L36K had stabilities of 4-5 kcal/mol at pH 7 the experiments with the less stable NVIAGLA and GLA backgrounds were possible.

The thermodynamic stability of five variants of SNase with the V23K substitution measured between pH 4 and pH 10 are shown in Figure 4.3. The marked decrease in stability with decreasing pH is consistent with a highly depressed pK_a for Lys-23. Lys-23 in the Δ+GLA background was previously studied and was determined to have a pK_a of 7.3 ± 0.1 [2]. The pH dependence of $\Delta G°_{H2O}$ for Δ+GLA/V23K showed evidence of two groups with depressed pK_a values. Fitting with Eq. 4.3 suggested the higher of the two pK_a values represented Lys-23; Asp-21 which normally titrates with a pK_a of 6.2 is suspected of being the other group with an altered pK_a in this protein[34]. A similar phenomenon was observed for Lys-23 in the more stable backgrounds, Δ+NVIAGA and Δ+NVIAGLA, consistent with more than one group with anomalous pK_a. Using Eq. 4.3, the pK_a of Lys-23 in Δ+NVIAGA and Δ+NVIAGLA were 7.3 ± 0.3 and 6.9 ± 0.3, respectively. Interestingly, in the two less stable backgrounds (NVIAGA and GLA), there is no evidence in the thermodynamic data suggesting the presence of multiple groups with abnormal pK_a values. Analysis with Eq. 4.2 yielded pK_a values of 5.7 ± 0.3 for Lys-23 in NVIAGA/V23K and 7.5 ± 0.3 in GLA/V23K (Fig. 4.5).

The thermodynamic stability of Lys-36 was also measured in the pH range 4 to 10 in all five background proteins (Fig. 4.4). As with the Lys-23 variants, the data from all variants was consistent with a depressed pK_a for Lys-36 (Fig. 4.5). Previous studies of Lys-36 in the Δ+GLA background reported a pK_a of 7.2 ± 0.2 [2]. Eq. 4.2 was used to fit the thermodynamic data for the more stable Δ+NVIAGA/L36K and Δ+NVIAGLA/L36K variants. The pK_a values resolved for Lys-36 in Δ+NVIAGA and Δ+NVIAGLA are 6.8 ± 0.2 and 5.3 ± 0.3, respectively. Both are more depressed than in the Δ+GLA variant. The thermodynamic data for NVIAGA/L36K showed evidence of more than one group with

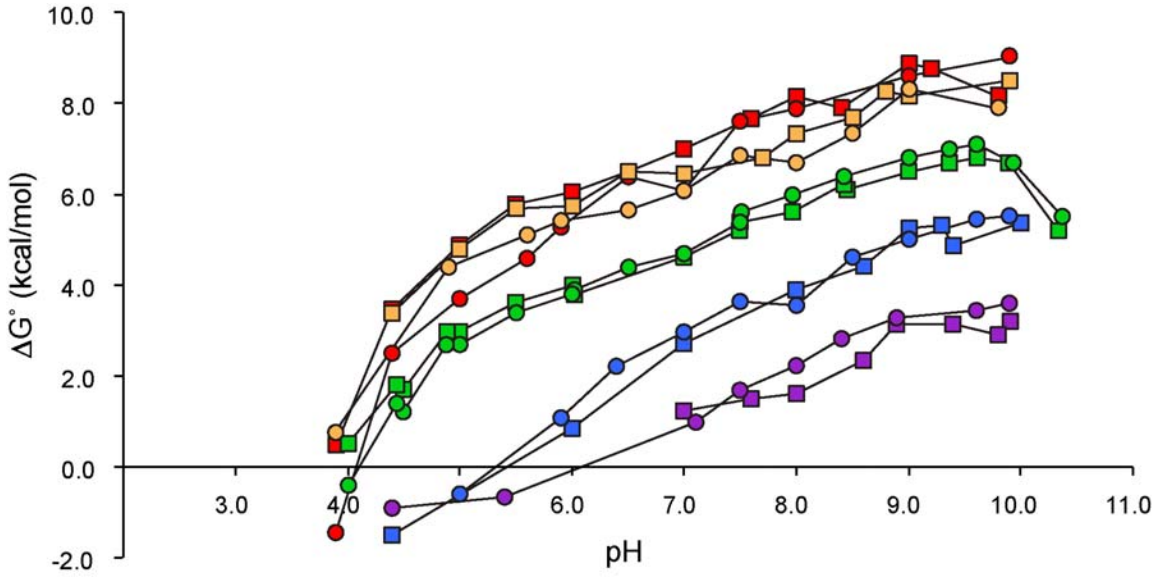

Figure 4.4 Thermodynamic stability ($\Delta G°_{H2O}$) of SNase variants as a function of pH. Data for proteins with (□) Lys-23 and (○) Lys-36, respectively. Colors identify the different SNase variants used as background for the substitutions with Lys-36: Δ+NVIAGLA (red), Δ+VIAGLA (orange), Δ+GLA (green), NVIAGA (blue), GLA (purple).

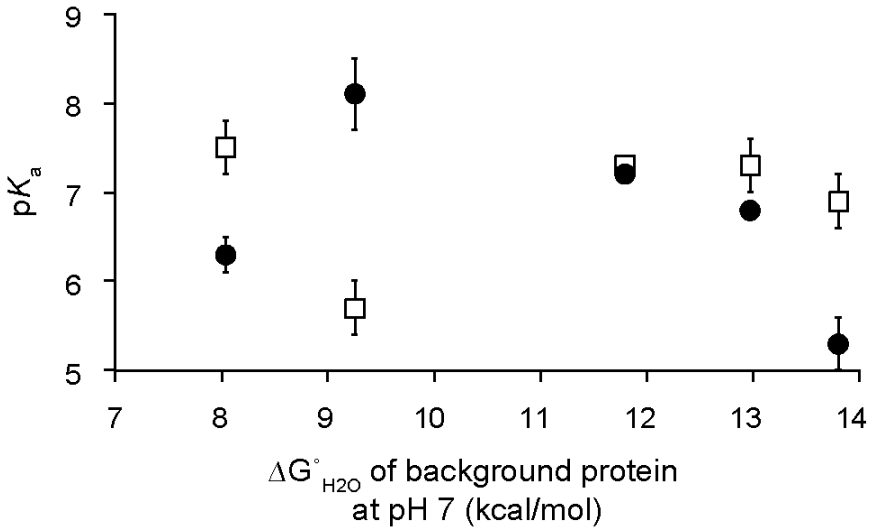

Figure 4.5 pK_a values of internal Lys compared to thermodynamic stability of the background protein. Data for proteins with (□) Lys-23 and (●) Lys-36, respectively. Error bars represent fit error from determining the pK_a values.

anomalous pK_a values, similar to that of Δ+GLA/L36K. Analysis with Eq. 4.3 provided a pK_a value of 8.1 ± 0.4 for Lys-36 in the NVIAGA background. Using Eq. 4.2, Lys-36 was determined to have a pK_a of 6.3 ± 0.2 in GLA, the least stable background. This is an unexpectedly lower pK_a for the Lys when compared with the observed trend in the more stable proteins.

4.4.2 CD spectroscopy

To determine if substitution of Lys-23 or Lys-36 affected the conformation of the SNase background proteins studied, far-UV CD spectra of each variant were collected at various pH conditions where, based on $\Delta G°_{H2O}$ measurements, the protein was expected to be fully folded (Fig. 4.6). In all backgrounds, the spectra for variants with Lys-23 or Lys-36 exhibited the signals at 208 nm and 222 nm characteristic of folded, native-like SNase. No significant changes in either signal intensity or shape of the spectra was observed for any of the 10 variants over the pH range 5 to 10. The absence of any large changes in the far-UV spectra as a function of pH suggests that the effects of Lys-23 or Lys-36 were insufficient to globally unfold the protein or to stabilize the protein in a very different conformation. This suggests that the structural reorganization coupled to the ionization of Lys-23 or Lys-36 is too subtle to be detected by this method.

4.4.3 Crystal structures

Crystal structures are available for three of the ten variants that were studied. Δ+GLA/V23K (PDB ID: 3QOJ) and Δ+GLA/L36K (PDB ID: 3EJI) have been refined to

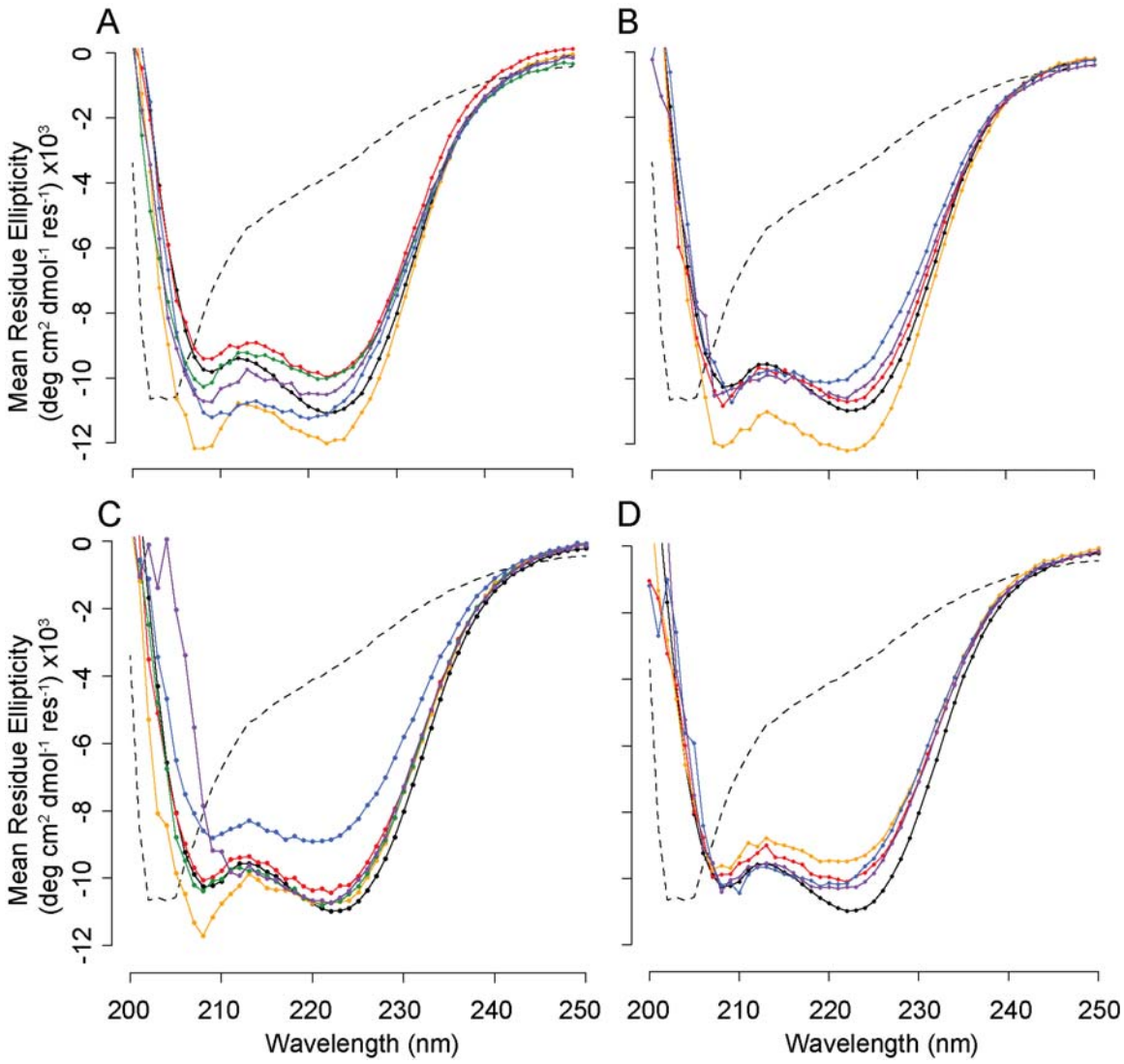

Figure 4.6 Far-UV CD spectra of Lys-23 and Lys-36 variants. (A) variants with Lys-23 at pH 7, **(B)** variants with Lys-23 at pH 10, **(C)** variants with Lys-36 at pH 7, **(D)** variants with Lys-36 at pH 10. Background SNase variants are displayed in color: Δ+NVIAGLA (red), Δ+VIAGLA (orange), Δ+GLA (green), NVIAGA (blue), GLA (purple).

1.60 Å and 1.75 Å, respectively, and were described previously[52]. Neither structure showed any evidence of a conformational difference to the parent proteins and have C_α RMS deviations < 0.4 Å when compared to the crystal structure of the background, Δ+GLA. The Lys side chain in both structures are completely buried in the hydrophobic core with no potential hydrogen bonding partners.

In contrast with what was observed in the Δ+GLA variants, the 1.75 Å structure of the Δ+VIAGLA/V23K protein exhibits a substantial reorganization (Fig. 4.7). Residues 15 to 24, comprising strands β1 and β2 of the five stranded β-barrel swing outward, exposing both Lys-23 and the entire the hydrophobic core to bulk solvent. In this open conformation, the Nζ atom of the Lys side chain is 2.3 Å away from the carboxylic group of Asp-19, which may be influencing the pK_a of Lys-23.

4.5 Discussion

A protein can respond to the presence of an internal Lys residue in a number of ways. If the buried Lys were surrounded by other polar groups from the protein, it could in principle be buried in the charged state without disrupting the structure by being stabilized by polar interactions. If the polar microenvironment were not as effective as water at solvating the charged Lys, the protein could respond by shifting into an alternative state in which the charged Lys could contact water. In an extreme case, the ionization of the buried group could be coupled to global unfolding. All instances of conformational reorganization would be driven by the need to solvate the charged Lys side chain with water or with an environment in the protein that is as good a solvent as water.

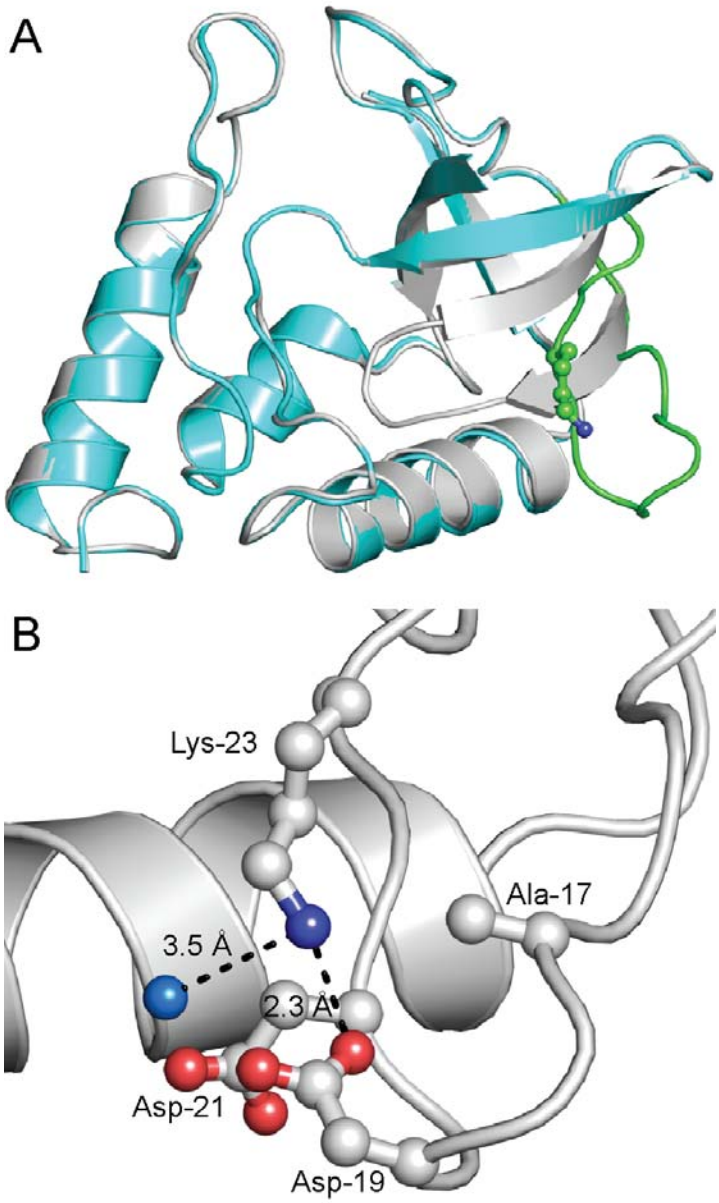

Figure 4.7 Crystal structure of Δ+VIAGLA/V23K. (A) Cartoon representation of the structure of Δ+GLA/V23K (white) overlaid with Δ+VIAGLA/V23K (cyan) with the conformational difference in β1-β2 conformational difference highlighted in yellow. The side chain of Lys-23 is shown as sticks. **(B)** Microenvironment of Lys-23 in Δ+VIAGLA/V23K showing all residues (sticks) and waters (spheres) visible within 5 Å of the Nζ atom of Lys-23.

The pK_a value measured with equilibrium thermodynamic experiments is a population weighted average of the pK_a of Lys in all the conformations that are accessed in solution[128]. To understand the origins of the anomalous pK_a values of buried groups it is of interest to understand the nature of the alternative conformational states that are reflected in the apparent pK_a. Secondary mutations that depress the global stability of the protein redistribute the population of the conformational ensemble. This would inherently affect the apparent pK_a values and afford a way of assessing the role of these alternative conformational states. As the global thermodynamic stability of the protein is decreased the probability increases that alternate conformational states are populated in which the internal Lys side chain is exposed to water. This would shift the apparent pK_a towards a more normal value.

4.5.1 Lys-23

The pK_a values of Lys-23 in the GLA, Δ+GLA, and Δ+VIAGLA background proteins ranged from 7.3 to 7.5 (Table 4.2 and Figure 4.8). This would appear to contradict the central hypothesis of this study. However, it is more likely that the pK_a of Lys-23 is independent of the thermodynamic stability of the background protein because it already reflects a substantial conformational reorganization that is sufficiently close in energy to the fully folded native state to render it insensitive to the effects of secondary mutations. This is suggested by the crystal structure of the Δ+VIAGLA/L23K protein showing that the β1-β2 strands are released from the barrel. This open state of SNase has been documented previously in SNase variants with the V23E substitution[58]. In this open state the ionizable side chains, Glu or Lys, are in bulk water. This open state is sufficiently

Table 4.2 pK_a values of Lys-23 and Lys-36 in 5 variants of SNase.

Protein	$\Delta G°_{H2O}$ [1]	Lys-23 pK_a [2]	Lys-36 pK_a [2]
GLA	8.0	7.5 (0.3)	6.3 (0.2)
NVIAGA	9.2	5.7 (0.3)	8.1 (0.4)
Δ+GLA	11.9	7.3 (0.1)	7.2 (0.1)
Δ+VIAGLA	13.0	7.3 (0.3)	6.8 (0.1)
Δ+NVIAGLA	13.8	6.9 (0.3)	5.3 (0.3)

[1] $\Delta G°_{H2O}$ measured in kcal/mol at 25℃ in 100 mM KCl and 25 mM HEPES at pH 7
[2] Fit-error of pK_a value shown in parentheses

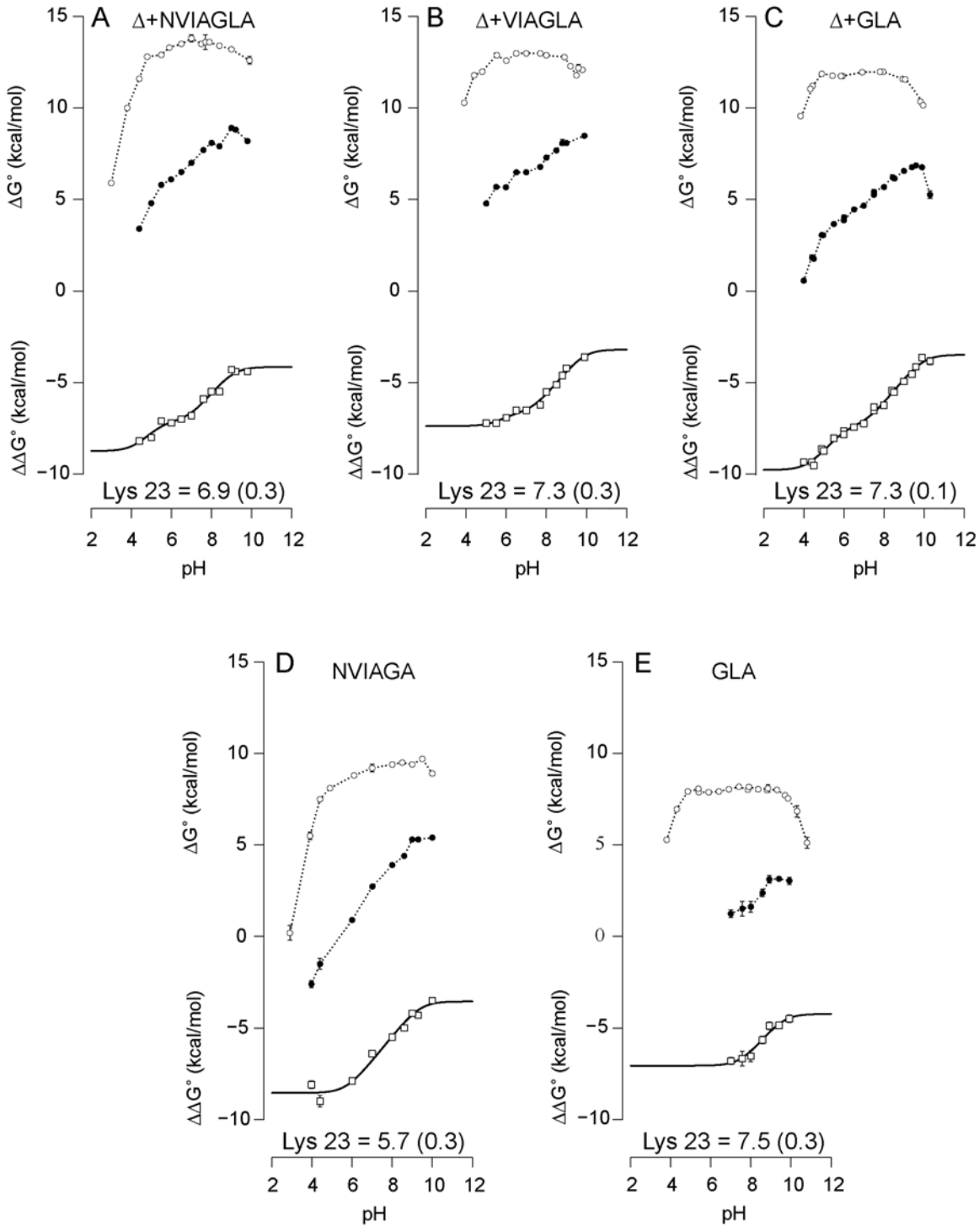

Figure 4.8 pH dependence of the stability of Lys-23 variants. pH dependence of stability of the background protein (○) and of the Lys-23 variant (●). Difference in stability ($\Delta\Delta G = \Delta G_{var} - \Delta G_{background}$) (□). Dotted lines are for guiding the eye. Black solid line describes the fit (Eq. 4.2 or 4.3) and the apparent pK_a of Lys-23 is at the bottom

close in energy to the fully folded state such that this state is easily populated regardless of the energy gap between the fully folded and fully unfolded states.

In contrast, Lys-23 in the NVIAGA/V23K and Δ+NVIAGLA/V23K proteins titrates with pK_a values of 5.7 ± 0.3 and 6.9 ± 0.3, respectively. These values are significantly different from the pK_a values of Lys-23 in the other backgrounds suggesting that the influence of stability on the pK_a is not as simple as previously hypothesized. It is possible that the combinations of mutations alter the conformational landscape of the protein to such a degree that the 'released' conformation with the opened β-barrel is inaccessible, leaving the pK_a of Lys-23 to be primarily determined by the protein transitioning from fully folded to fully unfolded. Another possible explanation involves the substitution of Asp-21, which has an elevated pK_a of 6.5, to Asn which is present in both NVIAGA and Δ+NVIAGLA[34]. It is reasonable that this mutation is too near to the Lys and is interacting in such a way as to depress the pK_a.

4.5.2 Lys-36

In contrast to the complex behavior of Lys-23, Lys-36 has a pK_a that is sensitive to the global stability of the protein. As the stability of the protein decreases from 13.8 kcal/mol in Δ+NVIAGLA to 9.3 in NVIAGA, the pK_a of Lys-36 shifts from a low value of 5.3 towards a higher, more normal value of 8.1 (Table 4.2 and Fig. 4.9). This is a substantial shift in pK_a. The interpretation of this shift is that in the less stable background protein an open conformation in which Lys-36 can access water is populated at a higher pH. It is known that mutations can alter the conformational distribution while

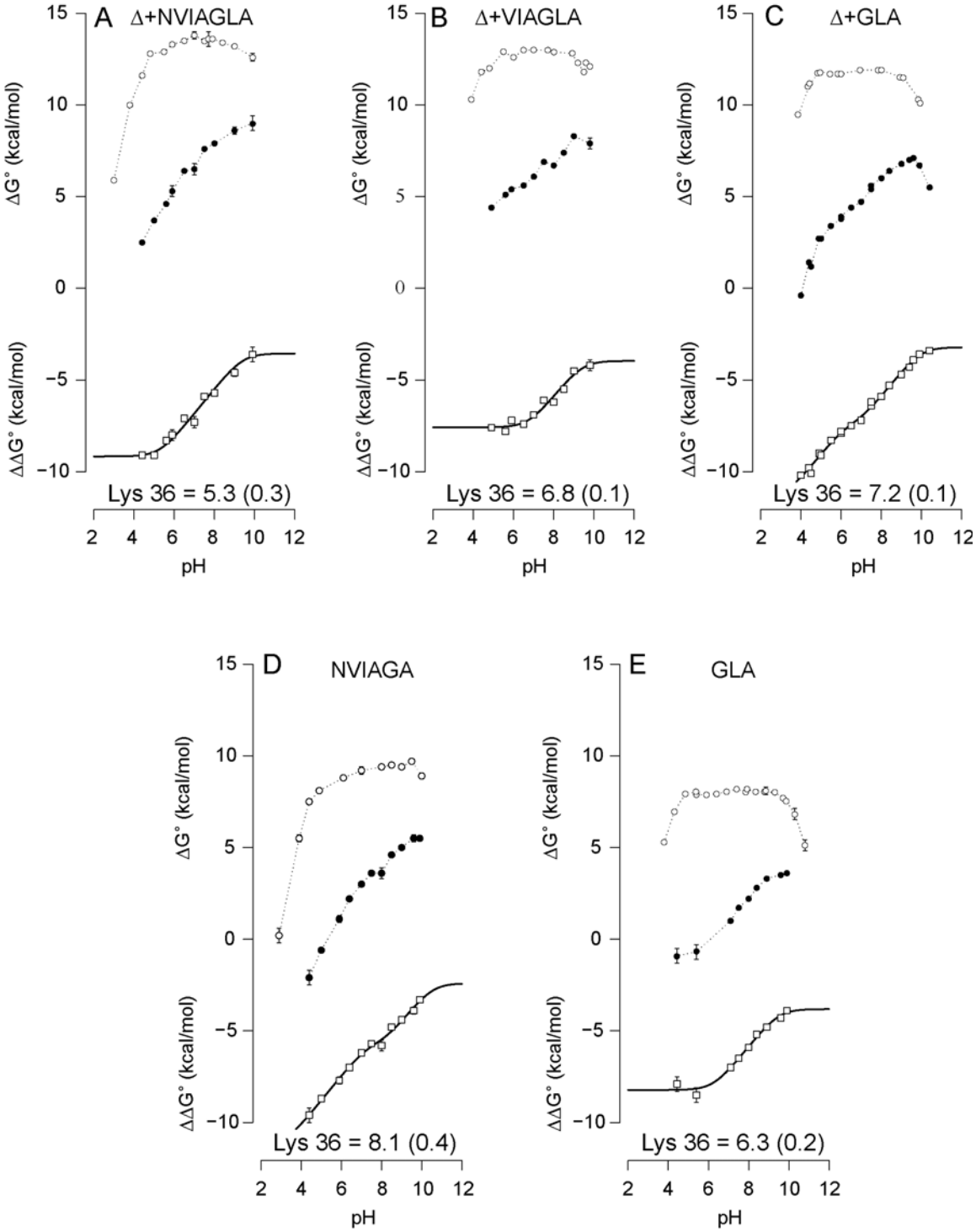

Figure 4.9 pH dependence of the stability of Lys-36 variants. pH dependence of stability of the background protein (○) and of the Lys-36 variant (●). Difference in stability ($\Delta\Delta G = \Delta G_{var} - \Delta G_{background}$) (□). Dotted lines are for guiding the eye. Black solid line describes the fit (Eq. 4.2 or 4.3) and the apparent pK_a of Lys-36 is at the bottom

not causing major structural changes[13]. The CD spectra of the SNase variants with Lys-36 did not reveal any large conformational changes. NMR spectroscopy studies will be useful to identify subtle conformational changes that might be responsible for the higher apparent pK_a in NVIAGA.

The pK_a of 6.3 of Lys-36 in GLA does not follow the pattern observed with the other proteins; it is too depressed. Again, this might be related to the effects of the NVIA substitutions on the conformational landscape. The pK_a values suggests that GLA/L36K cannot reorganize to hydrate the buried Lys-63. The reorganized state must be higher in energy than the unfolded state. This would cause the pK_a of Lys-36 to be solely dependent on the fully folded to fully unfolded populations, consistent with the data showing that the stability of GLA/L36K approaches 0 kcal/mol at pH of 6.

4.5.3 Implications about determinants of pK_a values of internal residues

It is known that the dynamic equilibrium between the different conformational states that constitute the energy landscape of a protein can be critical to function[118]. In the case of ionizable residues, their roles in biological energy transduction are often enabled by conformational reorganization which is why it is important to understand how the ionization of buried groups is coupled to reorganization and to shifts in the conformational ensemble. This can involve something as subtle as the reorganization of side chains and internal water molecules that enable proton pumping in bacteriorhodopsin[80,84,129] to the more dramatic coupling between side chain fluctuations and large amplitude motions as in the rotary motor in ATP synthase[19,85] or in the pH-driven regulation of capsid assemblies[130]. The demonstration that the anomalous pK_a

value of a buried ionizable group can be affected by the global thermodynamic stability of a protein implies that the magnitude of the shift in pK_a is governed partly by the probability of populating partial or locally unfolded conformational states, or a fully unfolded state in extreme cases, in which the previously buried ionizable group can be exposed to water.

The results of this study have important implications for structure-based pK_a calculations of buried ionizable groups. We have already demonstrated using NMR spectroscopy that the pK_a values of surface Glu and Asp residues in SNase are governed partly by the conformational flexibility of the backbone[53]. At pH values below the normal pK_a of 4 for Glu and Asp in water, the backbone can become increasingly disordered in a way that maximizes exposure of Glu and Asp side chains to water. This suggests that to predict the pK_a values of these residues, it is necessary to not only calculate the interaction of the ionizable side chains with other charged moieties and polar groups correctly. It is also necessary to reproduce the pH-dependence of the conformational fluctuations of the backbone so pK_a values can be computed in all the conformational states that contribute to the apparent pK_a values measured experimentally.

The case of buried ionizable groups appears to be more severe and the crystallographic data (Fig. 4.7) suggest that the conformational reorganization coupled to ionization of the buried groups can be more dramatic than what was observed with NMR spectroscopy for surface Asp and Glu residues. To predict the pK_a values of these buried residues, the alternative conformations have to be predicted correctly and the pH dependence of the population in the ensemble has to be captured accurately. That is, the probability that the alternative states are populated in solution at the pH condition of

interest has to be predicted which requires accurate calculation of Gibbs free energies which is a challenging proposition. Constant pH MD simulations attempt to explicitly reproduce the coupling between the change of ionization of a residue and the conformational reorganization associated with the change in charge[131–133]. Some computational models attempt to increase the accuracy of difficult cases of pK_a prediction by more deliberate sampling of alternative backbone conformations[134]. Some of the data in this study, especially outliers such as GLA/L36K and NVIAGA/V23K, will be especially useful to guide efforts to improve computational methods. Efforts are ongoing to elucidate structural details of the conformational changes that are coupled to the ionization of internal residues.

Chapter 5

Engineering of protein pH switches using buried ionizable groups with anomalous pK_a values

Done in collaboration with Peregrine Bell-Upp who contributed Δ+PHS/V66E/A109E data.

5.1 Abstract

Proteins capable of sensing and responding to small changes in pH in the physiological range are of significant biotechnological interest, especially in the area of cancer therapeutics. Many natural pH sensors involve His residues, but engineering artificial pH sensors with ionizable residues that normally titrate in the pH range of interest has turned out to be extremely challenging. Here, we introduce an approach for engineering pH sensing proteins based on the use of Lys or Glu residues that titrate with anomalous pK_a values as a result of being buried in the hydrophobic interior of a protein. In water, Lys and Glu titrate with pK_a values of 10.4 and 4.4, respectively, far from the physiological pH range. When buried in the hydrophobic interior the pK_a values of Lys and Glu residues can become highly depressed or elevated, respectively. These buried residues ionize in the physiological pH range triggering conformational reorganization. The utility of buried ionizable groups for design of artificial pH switches was illustrated in two variants of staphylococcal nuclease (SNase) with either the V66E/A109E or T62K/L125K substitutions. Far-UV CD and NMR spectroscopy along with equilibrium thermodynamic measurements were used to show that these two variants unfold globally and cooperatively in response to small changes in pH near pH 7. The variant with internal Lys residues is unfolded by a decrease in pH whereas the variant with the internal Glu residues unfolds in response to an increase pH. The exact range of pH where unfolding takes place is governed by the pK_a values of the internal Lys or Glu residues and by the global thermodynamic stability of the protein. This strategy for the design of artificial pH sensors is based on general and transferable thermodynamic principles and can be used to engineering pH sensitivity in the structure or function of any protein.

5.2 Introduction

Tight regulation of physiological pH is an essential organizing principle common to all living systems[164]. pH homeostasis is complex and involves the coupled regulation of chemical potentials of all ionic species, water, and osmolytes across bilayers. In eukaryotic cells, there are small, but meaningful, differences in the pH of the mitochondria (8.0), extracellular spaces (7.4), cytosol (7.2), endosomes (5.5-6.5), lysosomes, ER (7.2), nucleus (7.2) and the Golgi network (6.7-6.0)[146]. Nowhere is cellular pH constant; it changes as part of normal physiological processes. For example, acidification is a normal part of exercising muscle tissue under hypoxic conditions[165], the progression of cellular apoptosis[144], and endosomal and lysosomal maturation[146]. Cellular pH can also change as a result of trauma or disease, as in the case of acidification of ischemic tissue after stroke[166], or of cancerous solid tumors through the Warburg effect[167,168,153].

Given how tightly regulated intracellular pH is, it is not surprising that many proteins have evolved to interpret small changes in pH as important biochemical signals to modulate biological activity. For example, hemoglobin undergoes a conformational transition under acidic conditions in exercised muscle, lowering its binding affinity, thereby delivering oxygen where it is needed[169]. The haemagluttinin protein on the lipid bilayer of the influenza virus undergoes a global conformational transition when exposed to the acidic conditions of the endosome triggering fusion of the viral and endosomal bilayers, which is necessary for the release of the nucleocapsid into the cytoplasm[155,170]. Similar responses are present in flaviviruses[171], filoviruses[172], and in HIV[130]. pH is thought to play an important role in many aggregation disorders, including several

neurodegenerative diseases[149–151,173] and type-2 diabetes[174]. Although global understanding of the pH-sensing machinery in cells is lacking, the molecular mechanisms whereby proteins drive key biological processes initiated by changes in physiological pH are beginning to be understood.

There is considerable interest in the development of artificial pH sensing proteins for biotechnological applications[175,176]. Small, pH-dependent peptides have been developed for delivery of therapeutic agents into cells[156,177] and for targeting cancerous tumors for imaging[157] or immunotherapeutic purposes[59,178] pH-dependent fluorescent probes have been developed for quantitative determination of intracellular pH[179–188]. There is potential to modulate responses in cells using engineered alternate-frame folding[158,189] or allosteric coupling[190,191]. pH-switches may even be used to purify specific proteins of therapeutic interest[161] or modify them to improve efficacy[192–194,160]. Additionally, for many biotechnological uses, the translocation of a protein through a target cell's bilayer will require an unfolding event, furthering the need to exploit a pH-tunable conformational change.

The thermodynamic principles behind pH sensing are well established[162]. pH-driven conformational changes that depend on differential proton binding between two different conformational states. This requires that ionizable groups in the different conformational states titrate with different pK_a values. When the pH-sensing capacity of a protein is essential for biological function, the structural motif that acts as the pH sensor must be robust; pH sensors must be immune to spontaneous mutations. This might explain why, in the cases of hemoglobin, haemagluttinin, and other natural pH-sensing proteins, the pH-sensing motifs involve small contributions from many ionizable groups.

Most of these motifs involve multiple histidine residues because this groups normally titrates in the physiological range of pH.

Engineering novel pH-sensing proteins by emulating the naturally occurring pH sensors based on histidine residues has turned out to be extremely challenging and might be impossible[162]. To engineer a distributed network of ionizable residues capable of pH sensing requires either introducing new ionizable residues or manipulating the pK_a values of existing ones with arbitrary precision, a currently insurmountable challenge using computational or rational design methods. This problem is compounded by the need of novel switches to be highly cooperative; while natural pH switches have been optimized by evolution to coordinate many groups titrating at once[162], tuning the pK_a values of many histidines *de novo* to accomplish the same task is computationally impractical.

Here we introduce a novel approach for engineering pH sensing proteins that are active in physiological pH regimes by driving a conformational transition with large energetic contributions from a small number of Lys or Glu residues. The strategy is based on two observations: (1) Lys, Asp, and Glu buried in the hydrophobic interior of a protein have anomalous pK_a values shifted by as many as 5 units from the normal pK_a of 10.4, 4.5, and 4.0 of Lys, Glu, and Asp in water, respectively[35,2]. These anomalous pK_a values fall within the physiological pH range and are thus useful to sense changes in pH in this range. (2) The thermodynamic stability of proteins with ionizable groups with these highly anomalous pK_a is pH-sensitive[60]. A single buried Lys with a depressed pK_a changes the stability of the protein by 1.36 kcal/mol per pH unit away from the normal pK_a of 10.4 for Lys in water. The ionization of these buried groups is therefore sufficient to trigger local or global conformational transitions[35,2].

5.3 Materials and methods

5.3.1 Proteins

All variants were engineered with the Δ+PHS background of SNase using the QuickChange kits and purified following the protocol previously described[49,195].

5.3.2 pH titration monitored by Trp fluorescence and CD spectroscopy

Acid-base titrations monitored by changes in circular dichroism (CD) were performed with an Aviv CD spectrometer model 215 (Lakewood, NJ). Titrations that monitored Trp fluorescence were performed with an Aviv Automatic Titrating Fluorometer 105. All data were collected at 25℃ in 100 mM KCl following procedures published previously[35,2]. Each sample was prepared with a protein concentration of ~50 μg/mL with a buffer mixture consisting of 6.25 mM each of KAcetate, MES, Tris, and CHES in 100 mM KCl. Titrant was 0.3 N KOH or HCl. Midpoints of pH-driven transitions were obtained by non-linear least squares fits with a two state model.

5.3.3 Thermodynamic stability

Guanidine hydrochloride (GdnHCL) titrations were performed with an Aviv Automatic Titrating Fluorometer 105 to measure thermodynamic stability over a wide range of pH values. All data were collected at 25℃. Samples were prepared with a protein concentration of ~50 μg/mL in 100 mM KCl. The buffers varied based on the pH of the measurement: 25 mM CAPS, 25 mM CHES, or 25 mM HEPES were used for pH ranges 9.5-10, 8-9, and 7-7.5, respectively. The titrant was 6 M GdnHCl in the appropriate buffer. The titrations were performed as described previously[49,100].

5.3.4 NMR spectroscopy

^{1}H-^{15}N HSQC experiments were measured for the V66E/A109E and T62K/L125K variants. Following previously established protocols[34], samples were split in two, one for titrating with acid and one for base. Data were collected at 298 K on a Bruker Avance II-600 equipped with a cryoprobe. All spectra were processed with NMRPipe[98] and analyzed with Sparky[99].

5.3.5 X-ray crystallography

Crystals of the V66E/A109E variant were grown using hanging-drop vapor diffusion. Protein concentration was 8.0 mg/mL. Crystals were grown with a 2/1/1 molar ratio of $CaCl_2$ to 3'-5'-thymidine diphosphate to protein in 25 mM phosphate at 4℃ at pH 6.0 and 30% MPD. (Table A4). Crystals were flash cooled in liquid nitrogen. Diffraction data were collected using a Bruker Duo Apex diffractometer. Frames were processed using Bruker's software. Initial phases were obtained by molecular replacement methods in Phaser using the following search model: Δ+PHS (PDB id:3BDC) with solvent and heteroatoms removed, all B-factors set to 20.0 $Å^2$ and side chains truncated to Ala at the mutation site and disordered side chains. Iterative model building and refinement were performed in Coot and the Refmac 5 suite of CCP4. Refinements were performed until convergence of Rfactor and Rfree. TLS refinement was performed.

5.4 Results

5.4.1 Design goals

The starting protein was a highly stable variant of SNase that is insensitive to pH over a wide range of pH values. The goal was to alter this protein so it would undergo a cooperative transition between fully folded and fully unfolded states in response to a small change in pH in the physiological pH range. The approach involved use of multiple buried Lys or Glu residues that have anomalous pK_a value when the protein is in the fully folded state, and normal pK_a values when the protein is unfolded. The pK_a values are anomalous in the fully folded state because charges are incompatible with the hydrophobic and relatively dry interior of proteins; therefore, in the native state the equilibrium between charged and neutral forms of the ionizable moiety shift in favor of the neutral form.

5.4.2 Thermodynamic principles

The potential for ionizable groups to trigger a conformational transition in response to a change in pH is governed by two factors: (1) the net magnitude of differences in pK_a values of ionizable groups in the two different conformational states[162], and (2) the global thermodynamic stability of the protein. Simulations in Figure 5.1 show how a single Lys with pK_a depressed by 1, 2, or 3 pK_a units, or one, two or three Lys residues with a pK_a depressed by 1 unit affect the pH dependence of stability of a protein (Fig. 5.1B). The single Lys with a depressed pK_a decreases the stability of the protein by 1.36 kcal/mol per pH unit (at 298 K) in the range bracketed by the normal pK_a in the unfolded state and by the depressed pK_a in the native state (Fig. 5.1A). In contrast, the

presence of two or three Lys residues gives rise to a steeper sensitivity to changes in pH, equivalent to destabilization by 1.36, 2.72, or 4.08 kcal/mol per pH unit, respectively (Fig. 5.1B).

The simulations in Fig. 5.1A and 5.1B show how one, two or three ionizable groups with shifts in pK_a values of different magnitudes can affect the stability of a protein. Whether these groups poise the protein for structural response depends on the balance between the free energy stored in the form of differences in pK_a values and the component of the net difference in the Gibbs free energy ($\Delta\Delta G°_{H2O}$ = $\Delta G°_{H2O}$ (folded) - $\Delta G°_{H2O}$ (unfolded)) that is insensitive to pH. At the pH where the stability lost by shifting of pK_a values is equal to the inherent, pH-independent component of stability, the folded and unfolded states of the protein will exist in equal amounts (i.e the $\Delta G°_{H2O}$ = 0 point).

5.4.3 Selection of pH-sensing moieties

The internal Lys and Glu residues studied previously in SNase were engineered into a form of SNase made highly stable ($\Delta G°_{H2O}$ = 11.8 kcal/mol at 298 K) with truncations and substitutions (del 44-49, G50F, V51N, P117G, H124L, S128A). With the exceptions of Lys-92, Lys-104, Lys-100, and Glu-92 no single internal Lys or Glu was sufficient to unfold SNase globally[35,2]. Simulations based on the experimentally determined thermodynamic stability of the variants and the measured pK_a values were used to identify pairs of Lys or Glu residues likely to trigger global unfolding in the neighborhood of pH 7.4, equivalent to the pH of human blood[146]. In the face of evidence to the contrary, additivity of thermodynamic effects was assumed although this is clearly

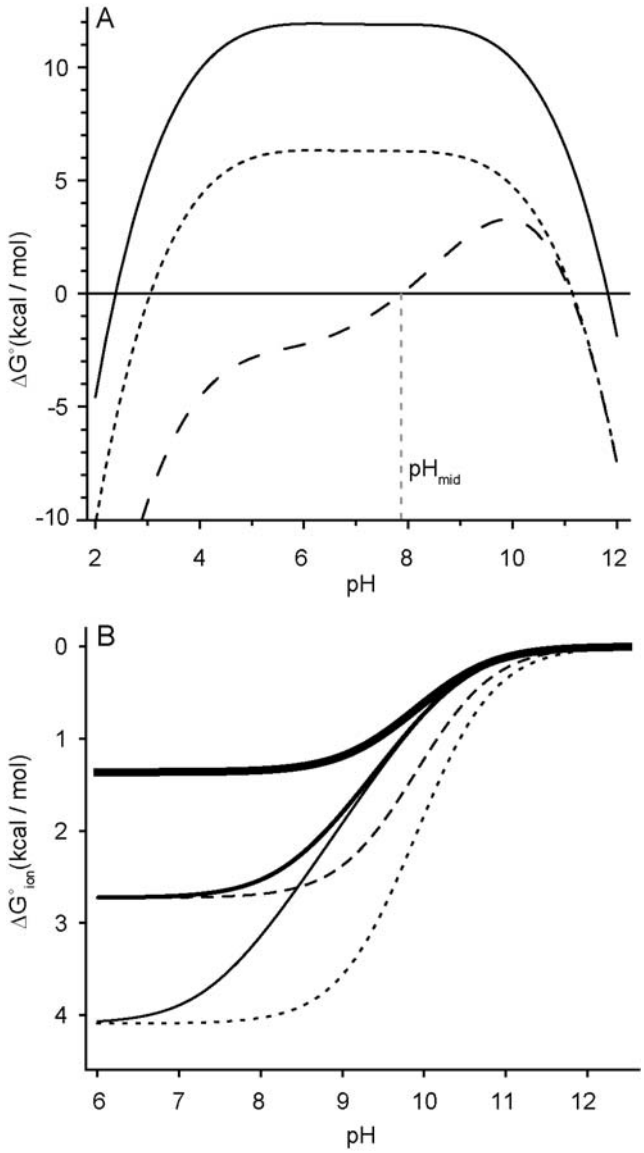

Figure 5.1 Simulations of the effects of buried ionizable residues with anomalous pK_a values on the pH dependence of thermodynamic stability of proteins. This case represents a protein with internal Lys residues. **(A)** pH dependence of stability of the background protein, the Δ+PHS variant of SNase (solid). Variant in which the buried Lys has a normal pK_a of 10.4 (dashed) and an anomalous pK_a of 7 (dotted). **(B)** Consequences of the insertion of internal Lys residue(s) on the stability of a protein relative to stability at high pH. Case of single Lys with a pK_a shift of 1, 2, or 3 pH units (solid thick, medium, thin lines, respectively) and 2 or 3 Lys with pK_a shift of 2 pH units (dashed, dotted, respectively).

an approximation unlikely to hold in all cases precisely because burial an ionization of internal residues are coupled to conformational reorganization.

Based on the simulations two variants were selected for detailed study: the variant with T62K/L125K and the variant with V66E/A109E. The crystal structure of the V66E/A109E protein was obtained but the T62K/L125K protein resisted crystallization (Fig. 5.2A and Table A4). Crystal structures of the individual Lys mutations (PDB id: 3DMU, 5C1E) show almost no deviations from Δ+PHS. The double Glu variant shows that the Glu-66 is fully buried and Glu-109 has slight accessibility to solvent-accessible, up to 5 $Å^2$ accessible to a crystallographic water (Fig. 5.2B). The structure of the double variant is almost identical to that of the background Δ+PHS protein with the exception of the loop region comprising residues 41 to 53. Whereas the crystal structure of the single variant, V66E (PDB id: 5EGT), is nearly identical to Δ+PHS, this same deviation in the loop region is seen in the single variant, A109E (PDB id: 4YIJ).

5.4.4 pH dependence of stability

The T62K/L125K protein was designed to be stable at high pH, where the Lys residues are neutral, and to unfold near pH 7.4, where normally they would be charged. In contrast, the V66E/A109E variant was designed to be stable at low pH, where carboxylic groups are usually neutral, and to unfold at higher pH, where carboxylic groups are normally charged. This behavior was demonstrated by measuring thermodynamic stability ($\Delta G°_{H2O}$) as a function of pH using chemical denaturation monitored by Trp fluorescence (Fig. 5.3A, Table 5.1). The background protein, referred to as Δ+PHS, has little to no pH sensitivity between pH 5 and 9 so any change in pH

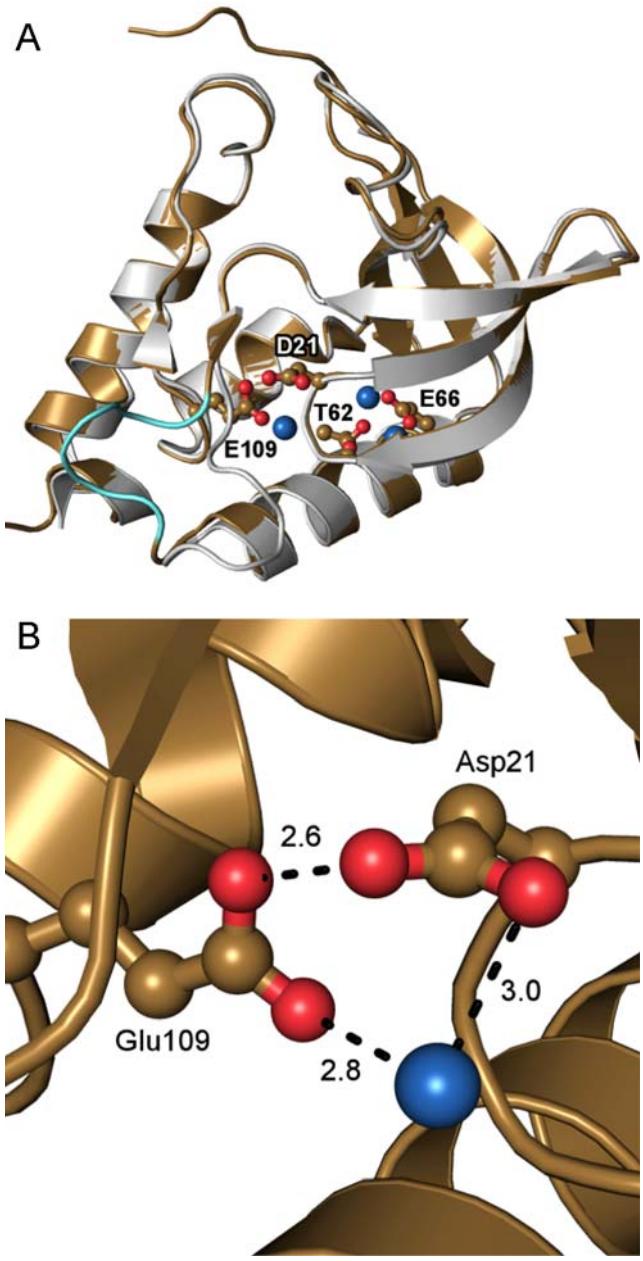

Figure 5.2 Crystal structure of the Δ+PHS/V66E/A109E variant of SNase. (A) Overlay of the structures of the background protein, Δ+PHS (gray) and the Δ+PHS/V66E/A109E variant (gold) (PDB ID 4OL7). Side chains of Glu-66 and Glu-109 are shown, as well as side chains in contact with the internal Glu residues. Large differences in the conformation of the backbone are highlighted in cyan. Crystallographic waters that coordinate internal Glu residues are represented in blue. **(B)** Microenvironment of the Glu-109 side chain showing close proximity to surface group Asp-21, and coordination of crystallographic water molecule, represented in blue.

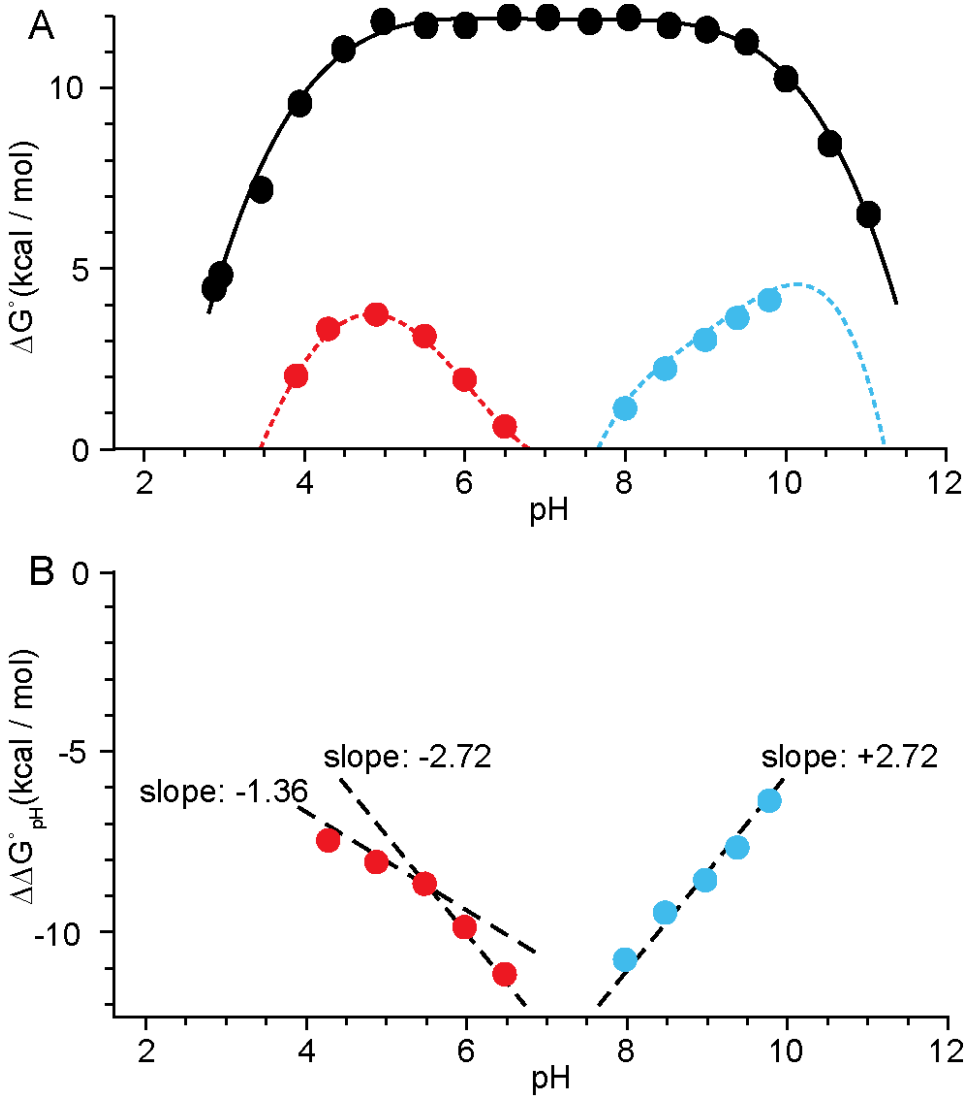

Figure 5.3 pH dependence of thermodynamic stability ($\Delta G°_{H2O}$) and difference in stability ($\Delta\Delta G°_{H2O}$) of pH switch proteins. (A) pH dependence of thermodynamic stability ($\Delta G°_{H2O}$) of the Δ+PHS form of nuclease (black), Δ+PHS/T62K/L125K (red) and Δ+PHS/V66E/A109E (blue) variants. Solid lines are meant only to guide the eye. **(B)** Difference in stability, calculated as $\Delta\Delta G° = \Delta G°(\text{variant}) - \Delta G°(\Delta\text{+PHS})$. Dashed lines representing ideal, expected behavior from the titration of 1 or 2 groups with anomalous pK_a, with slopes of 1.36 or 2.72, respectively.

Table 5.1 Thermodynamic parameters of Δ+PHS/T62K/L125K and Δ+PHS/V66E/A109E

Protein	pH	$\Delta G°_{H2O}$ (kcal/mol)	Cm	m
Δ+PHS	9.9	10.3 ± 0.1		
	9.5	11.3 ± 0.1		
	9.0	11.5 ± 0.1		
	8.5	11.6 ± 0.1		
	8.0	11.9 ± 0.1		
	7.0	11.9 ± 0.1		
	6.5	11.9 ± 0.1		
	6.0	11.7 ± 0.1		
	5.5	11.7 ± 0.1		
	4.9	11.8 ± 0.2		
	4.4	11.0 ± 0.2		
	3.9	9.5 ± 0.2		
Δ+PHS/T62K/L125K	9.8	4.1 ± 0.1	0.7	6.1 ± 0.1
	9.4	3.6 ± 0.4	0.6	6.1 ± 0.4
	9.0	3.0 ± 0.3	0.5	6.2 ± 0.4
	8.5	2.2 ± 0.4	0.3	6.2 ± 0.5
	8.0	1.1 ± 0.5	0.2	6.0 ± 0.8
Δ+PHS/V66E/A109E	6.5	0.6 ± 0.6	0.2	5.9 ± 1.2
	6.0	1.9 ± 0.5	0.4	5.7 ± 0.7
	5.5	3.1 ± 0.2	0.5	6.0 ± 0.2
	4.9	3.7 ± 0.3	0.7	5.6 ± 0.1
	4.3	3.3 ± 0.2	0.6	4.9 ± 0.1
	3.9	2.0 ± 0.1	0.4	4.4 ± 0.1

sensitivity in this region can be attributed to the internal Glu or Lys residues. The consequences of the internal Lys or Glu residues on the pH sensitivity of the protein are fully consistent with what observed previously[61,62,63]. For both proteins $\Delta G°_{H2O}$ approaches 0 near pH 7, consistent with the idea that the proteins will be unfolded near neutral pH.

Close examination of the dependence of $\Delta G°_{H2O}$ on pH ($\Delta\Delta G°_{H2O}$ / pH where $\Delta\Delta G°_{H2O} = \Delta G°_{H2O}$ (variant) - $\Delta G°_{H2O}$ (background)) was used to examine if one or both ionizable groups play a role in determining the pH sensitivity of the variants. Based on previous studies from this laboratory, it is reasonable to assume that the pH sensitivity the variants can be largely attributed to the internal Glu and Lys residues. The rate of change of $\Delta G°_{H2O}$ with pH for every shift in pK_a is 1.36 kcal/mol for a single titrating group, or 2.72 kcal/mol for two independently titrating groups. If both Glu-66 and Glu -109 or Lys-62 and Lys-125 have shifted pK_a values, the data in Fig. 5.1B would fit well with the plotted ideal slopes of ± 2.72 kcal/mol/pH. This is the case for the T62K/L125K protein. The V66E/A109E variant fits a 2-group slope at pH > 5.5 but is closer to a 1-group slope at pH < 5.5 indicating that the interactions between E109 and D21 observed in the crystal structure (Fig. 5.2B) affects the pK_a values of these groups and that one of these two groups has a more normal pK_a than when the group exists in the presence of the other. Based on the crystal structure, it is reasonable that Glu-109 has a more normal pK_a than Glu-66.

5.4.5 pH switching behavior

Trp fluorescence, far-UV circular dichroism (222 nm), and NMR spectroscopy showed that the two proteins switch cooperatively between folded and unfolded states near pH 7.4 and in response to changes in pH.

The region of interest in the acid/base titrations by Trp fluorescence (Fig. 5.4A) and far UV-CD (Fig. 5.4B) is the one centered near pH 7. The acid titration observed for the double Glu variant (red) is of no special interest; it is shifted to higher pH relative to the background protein (black) because its thermodynamic stability is decreased. The midpoints of the unfolding of the double Lys variants monitored by Trp fluorescence and CD 222 nm were 7.5 and 7.4, respectively. For the double Glu variant they were 6.8 and 6.9, respectively (Table 5.2). The pH switch encoded by the Glu residues is more sensitive than the

one encoded by Lys residues, thus the transition between folded and unfolded states is steeper.

The double Lys variant exhibited a secondary transition at pH < 6, more clearly apparent in the CD spectroscopy data (Fig. 5.4B). The baselines for the double Glu variant at pH > 7 suggest that the unfolded states populated by the two different proteins near pH 7 are structurally different. Full CD scans in the far-UV range support the notion that the base unfolded form of the double Glu variant is more structured than the unfolded form of the double Lys variant (Fig. 5.5). Indeed the CD data are consistent with the loss of secondary structure relative to fully folded and unfolded SNase. The spectrum of the V66E/A109E protein at pH 5, where it is most stable, has minima near 208 nm and 222 nm like that of the fully folded reference spectrum but the ratio is

Table 5.2 pH titrations of Δ+PHS/T62K/L125K and Δ+PHS/V66E/A109E.

Protein	(predicted) pH_{mid}	Signal	(major) pH_{mid}	(minor) pH_{mid}
Δ+PHS		CD (222 nm)	2.06 ± 0.02	-
		Fluorescence	2.03 ± 0.08	4.39 ± 0.41
Δ+PHS/T62K/L125K	8.0	CD (222 nm)	7.43 ± 0.03	3.35 ± 0.31[a]
		Fluorescence	7.54 ± 0.05	-
Δ+PHS/V66E/A109E	5.9	CD (222 nm)	6.75 ± 0.01	-
		Fluorescence	6.92 ± 0.02	-

[a] This transition represents the acid-unfolding of the protein

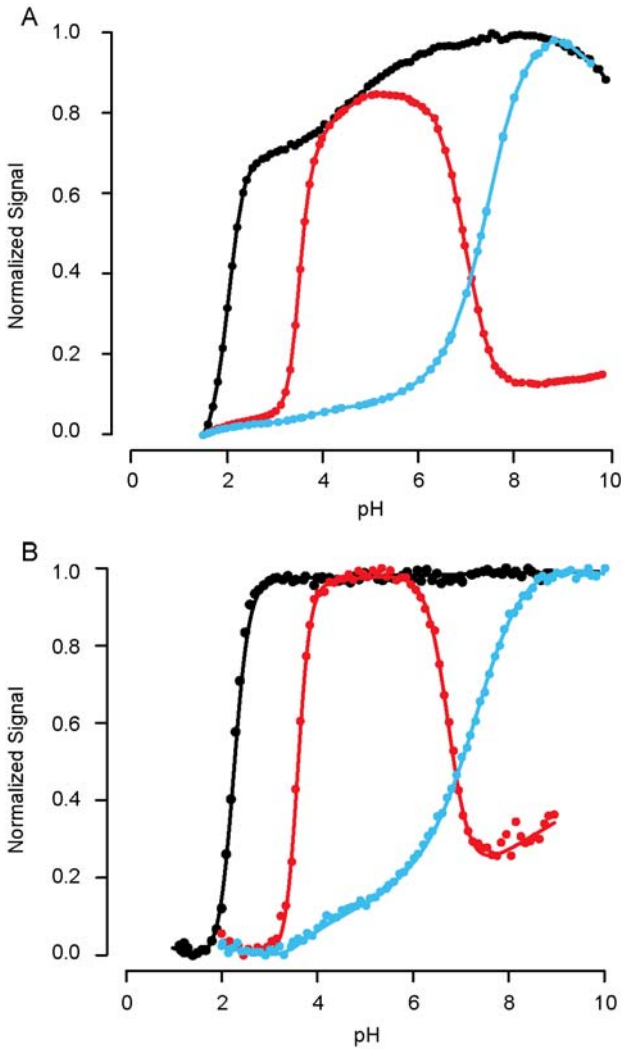

Figure 5.4. pH titrations of switch proteins. Acid titrations monitored by **(A)** Trp fluorescence and **(B)** CD at 222 nm for Δ+PHS (black), Δ+PHS/V66E/A109E (red), and Δ+PHS/T62K/L125K (blue). Data have been normalized relative to the Δ+PHS curves.

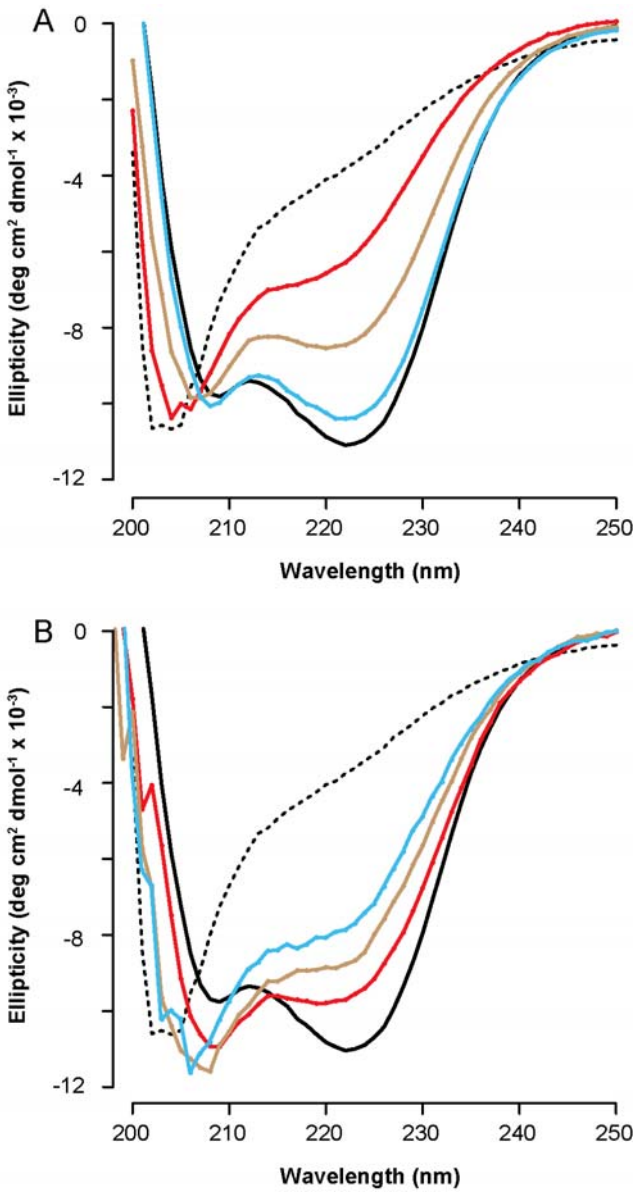

Figure 5.5 Far-UV CD spectroscopy of switch proteins. Far-UV wavelength scans for **(A)** Δ+PHS/T62K/L125K and **(B)** Δ+PHS/V66E/A109E at pH 5 (red), 9 (blue), and 7.5 or 7 (brown) for (A) and (B), respectively. Black curves are representative of folded (solid) and unfolded (dashed) SNase at pH 7.

reversed; the signal at 208 nm is stronger than at 222 nm. As the pH increases, there is a further shift in the ratio of minima, which is expected of a protein that has lost secondary structure. The spectra of the T62K/L125K protein show the same shift in the ratio 208 nm to 222 nm when the pH decreased, as expected. The spectrum at pH 9 closely matches that of the folded state while the spectrum at pH 5 appears similar to that of the unfolded state.

The agreement between the structural transitions reported by fluorescence and by CD was excellent. ^{1}H-^{15}N HSQC spectra collected at pH 5.09 and 7.94 for the V66E/A109E protein and at pH 6.49 and 8.53 for the T62K/L125K protein were used to obtained a more atomistic level of the structural changes involved in the pH switch (Fig. 5.6). At the higher pH, the spectrum for the T62K/L125K variant is well dispersed and matches closely the spectrum of the background protein (data not shown). At low pH where the protein was expected to be unfolded, all resonances collapse into the pattern without dispersion, characteristic of an unfolded protein. The case of the V66E/A109E protein is more complex. At pH 5.09 where the protein as expected to be fully folded, the spectrum shows peaks consistent with a fully folded population alongside peaks characteristic of unfolded protein. On the other hand, at pH 7.94 where the protein is expected to be fully unfolded, the spectrum was fully consistent with that of a fully unfolded protein.

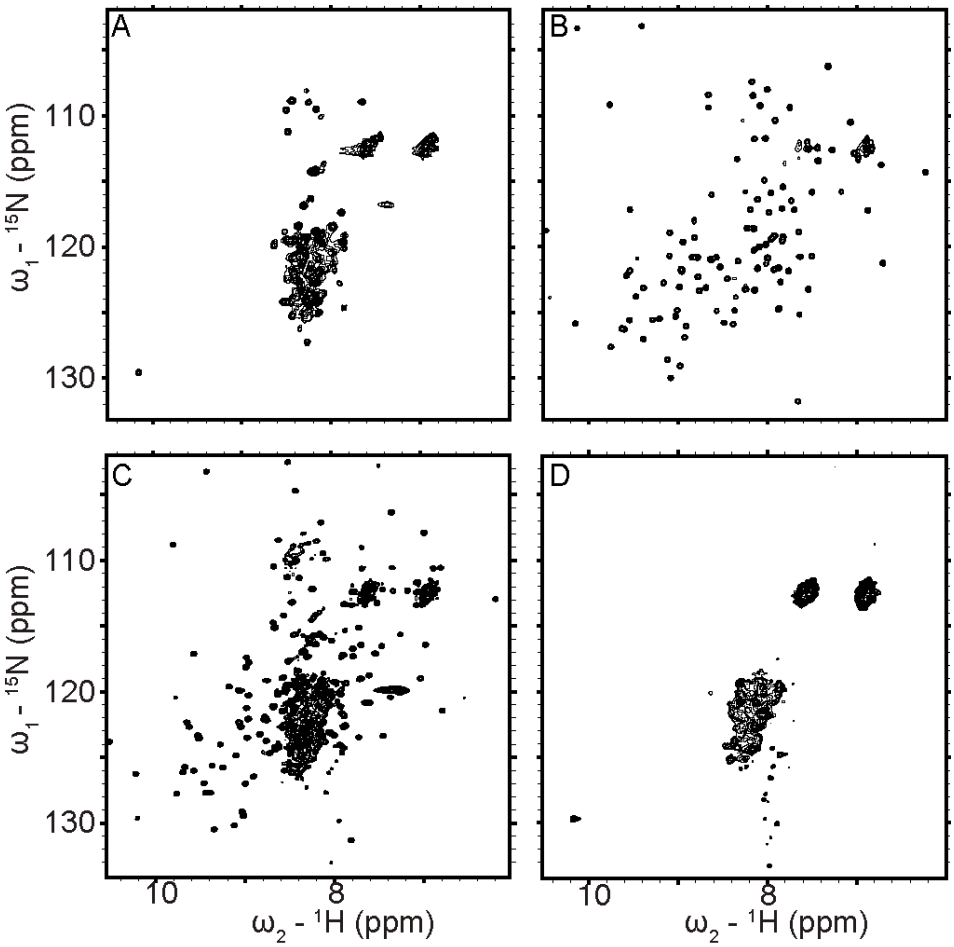

Figure 5.6 NMR spectroscopy of Δ+PHS/T62K/L125K and Δ+PHS/V66E/A109E. HSQC spectra of Δ+PHS/T62K/L125K at pH 6.49 (a) and pH 8.53 (b) and Δ+PHS/V66E/A109E at pH 5.09 (c) and pH 7.94 (d)

5.5 Discussion

Based on previous measurements demonstrating that buried ionizable groups titrate with anomalous pK_a values, many close to 7, and on fundamental principles of linkage thermodynamics, we successfully engineered two different kinds of robust pH switch proteins that are activated by changes in pH in the physiological range. We have demonstrated that buried ionizable groups with anomalous pK_a values are useful sensors of pH and can be used to engineer pH switches capable of undergoing large, highly cooperative conformational transitions between fully folded and unfolded states in response to small changes in pH in the physiological range.

The Δ+PHS form of SNase used in all our previous studies[1,2,35,60] is too stable to be unfolded by the ionization of a single residue, but simultaneous burial of two ionizable groups with anomalous pK_a was sufficient to turn the variants into pH-sensitive switches that operate near pH 7. The residues that were substituted by Lys or Glu were selected partly based on their locations, on the measured pK_a values when buried singly, and on the thermodynamic consequences of the substitutions[60]. Because each Lys or Glu experiences a substantial shift in pK_a relative to normal pK_a values in water, this approach avoids the necessity of having to make many substitutions to engineer the pH sensitivity of the switch.

The predicted pH_{mid} values based on our simulations for the T62K/L125K and V66E/A109E proteins were 8.0 and 5.9 respectively. They were within 1 pH unit of the measured values (Table 5.1), suggesting that, at least in the case of these ionizable groups, although additivity is not perfect, information about the single buried ionizable groups is sufficient to predict the range of pH where the switch will be active.

The potential of buried Lys and Glu residues to act as pH sensors is inherently different. The Lys side chain is longer and more flexible and could, in principle, be more difficult to bury than the shorter Glu side chain. On the other hand, the charge in the Glu side chain is delocalized, raising the possibility that burial of Glu residues is tolerated better than Lys. The properties of buried Lys and Glu and even Asp residues as pH sensors need to be examined further.

The data show that the switch proteins do switch between mostly folded and mostly unfolded states. One of the potential problems that could have been encountered is with lack of cooperativity in the structural transition of interest. The protein could have resolved the electrostatic crisis originating with burial of an ionizable group in a hydrophobic environment with a local conformational reorganization, as suggested by NMR spectroscopy studies[52]. Hysteresis could have been an issue but this was ruled by fluorescence behaviors showing that the behavior and pH_{mid} values from acid or base titrations were the same (data not shown).

The degree of pH-driven unfolding of the two types of switch proteins was not the same. The CD spectroscopy data for the T62K/L125K variant suggests that this protein populates an intermediate state between the folded and the acid unfolded state. This state, however, is not observed in the HSQC data because at low pH the HSQC spectra report on the unfolded state because it is the dominant population, whereas the CD and fluorescence experiments report primarily on the folded population. Overall the data show that at high pH the folded state is the dominant state and that at low pH the dominant state is the unfolded one.

The case of the V66E/A109E protein is more complex. The crystal structure (Fig. 5.2A) shows that except for rearrangement of a loop, the structure of the double Glu variant is comparable to the fully folded parent protein. However, in pH titrations monitored by CD and fluorescence, the acid and base denatured baselines do not reach the same signal levels. The free energy gap between the folded and base-unfolded states might be small enough to allow sporadic fluctuations to the folded state. This is consistent with the NMR data showing a predominantly unfolded population while the CD and fluorescence data betray the presence of folded protein even at high pH. The HSQC spectrum of the V66E/A109E protein at pH 5 suggests that populations of folded and partially or fully-unfolded protein co-exist. The V66E and A109E substitutions may have triggered local unfolding or destabilization beyond what is detectable in a crystal structure. This is consistent with fact that the CD wavelength scan at pH 5 does not resemble either the folded or the unfolded standard. This does not run contrary to the acid-base titration experiments as those depend on a Trp or some amount of secondary structure to be present. Overall the data suggest that the T62K/L125K variant switches between fully folded and fully unfolded better than the V66E/A109E variant. The reasons behind these effects will be examined by engineering many other pH switches with double Lys and double Glu substitutions in SNase.

The approach for the design of these pH switch proteins is based on general physical properties of proteins and on general thermodynamic principles. The approach should be general and transferable to any other protein.

Chapter 6

Rational engineering of pH switch proteins using buried Lys residues

6.1 Abstract

The ability to alter proteins rationally to render them sensitive to small changes in pH in physiological pH regimes is highly desirable for many biotechnological applications. Natural pH switches usually depend on His residues with pK_a values near physiological pH but rational design of pH switches based on residues that normally titrate in the pH range of interest has proven to be very challenging. Previously we showed that staphylococcal nuclease can be converted into a pH switch simply by substituting internal positions 62 and 125 with Lys. This artificial pH sensor responds to small changes in pH near pH 7 with a highly cooperative conformational transition. Now we have studied twenty-five variants of staphylococcal nuclease, each with two internal Lys residues, to demonstrate that this is a robust design principle for the engineering of pH switch proteins active in the physiological range of pH. The internal Lys residues that were used titrate with pK_a values that are depressed relative to the normal pK_a of 10.4 for Lys in water. Trp fluorescence and CD spectroscopy were used to demonstrate that the majority of these proteins act as pH switches that cooperatively unfolding in response to small changes in pH at or near physiological values. The pH dependence of stability was also measured to characterize the H^+ binding reaction. Crystal structures suggest that when the Lys residues are buried and neutral the structure of the protein is fully folded and very similar to the parent protein. As the pH approaches the pH of unfolding, local structural rearrangements are observable. This design principle for the engineering of pH switches with buried Lys residues with depressed pK_a values is based on fundamental thermodynamic principles; it is general and transferable to any protein.

6.2 Introduction

pH is very tightly regulated in all living organisms. Metabolic processes[135–137], proton/ion-coupled transport[138,139], and H^+-ATPases[140,141] work in concert to achieve and maintain specific distribution of pH values in different parts of the organism. Cell growth[142], motility[143], apoptosis[144] and many other cellular processes are highly sensitive to pH. Pathological conditions such as osteoporosis[145,146], cardiac arrhythmias or failure[147,148], and some protein aggregation diseases[149–151] are also known to be influenced by transient or permanent alterations in pH homeostasis. The pH is slightly different in different parts of the cell. For example, the cytoplasm has a pH near 7.2, while mitochondria are more basic and endosomes are more acidic than the cytoplasm[146]. Dysregulation of pH in solid cancerous tumors constitutes a striking example of the importance of proper pH regulation[152–154]. Despite recognition that changes in pH act as biological signals, that proteins serve as the receptors for these signals, and that dysregulation is central to progress of many pathological states, the manner in which H^+ act as a biological messenger is only now beginning to be examined systematically.

Many proteins have evolved to rely on small changes in pH as a signal to modulate biological activity. In some of these proteins, changes in pH influence function by triggering conformational transitions. For example, the haemagluttinin protein of the influenza virus undergoes a global conformational reorganization when the virus is exposed to the acidic conditions of the endosome. This conformational change triggers fusion of the viral and endosomal bilayers necessary for the release of the nucleocapsid into the cytoplasm[155].

The increased recognition of the role of pH dysregulation in pathological

conditions such as cancer has led to attempts to exploit the characteristic pH profile of cancerous tumors for therapeutic or diagnostic purposes. One of the hallmarks of cancer is the acidification of the extracellular environment. Small peptides have been developed to transport genes or small molecules into cells as a delivery mechanism for gene therapy[156] or for cell regulation and diagnostic imaging of cancerous tumors[157]. Scanning mutagenesis has been employed in attempts to engineer pH sensitive antibodies that can exploit the pH differences between cancerous and normal tissues[59]. pH switch proteins have also been designed by either combining proteins with pre-existing switch-like behavior[158,159] or introducing mutations to a previously pH-insensitive protein[59,160,161].

The ability of a protein to act as a pH sensor that responds to a small change in pH depends on the differential H^+ binding affinities of ionizable residues between different protein conformations and on the free energy difference between these conformational states[162]. The ionizable residues responsible for pH sensitivity must titrate with different pK_a values in the different conformational states of the protein. Natural pH sensors tend to distribute the pH sensing activity into small contributions from many ionizable residues. This ensures a robust response insensitive to spontaneous mutations. Emulating this distributed pH sensor mechanism in the lab is very challenging because every attempt to alter the pK_a value of a residue also alters the thermodynamic stability of the protein, which is one of the determinants of pH sensing activity[162]. In this study we test a design principle for the engineering of pH sensing proteins that depends on harnessing very large contributions from a small number of ionizable residues with highly anomalous pK_a values.

A preliminary set of experiments showed that staphylococcal nuclease (SNase)

could be turned into a pH sensor that responds with global unfolding near pH 7, simply by burial of either two Lys residues or two Glu residues (Chapter 5). These buried ionizable residues had anomalous pK_a values, depressed for the Lys residues and elevated for the Glu residues. These shifts in pK_a render the stability of the protein highly pH sensitive. In the protein with two internal Lys residues, the stability decreases with decreasing pH, and in the one with two Glu residues, it decreases with increasing pH. Stability is lost at the rate of approximately 1.36 kcal/mol per pH unit for every buried ionizable group with a shifted pK_a [2]. The two proteins that were studied undergo highly cooperative unfolding transitions near pH 7 driven by the ionization of the two buried groups.

The pK_a values of 25 internal Lys residues in SNase have been measured previously[2]. The pK_a values of most of these Lys residues are depressed below the normal pK_a of 10.4 of Lys in water. These shifts in pK_a represent a substantial thermodynamic force capable of driving local or global conformational changes[2,35,123]. In this study, we took advantage of the existing database of Lys residues with anomalous pK_a values to show that many different pairs of buried Lys residues could be used to turn SNase into a pH switch. The goal was to demonstrate the robustness of this design principle and to identify cases where it fails.

A set of 25 variants of SNase was engineered using pairs of internal Lys residues. The Lys residues were introduced in different elements of secondary structure. The distances between the two Lys residues was varied to determine the extent to which this affected the cooperativity of the unfolding of the protein. Trp fluorescence and CD spectroscopy were used to monitor the structure of the protein in pH titrations and to

measure the thermodynamic stability so we could establish if one or both buried Lys residues were involved in H^+ binding upon unfolding. Thermodynamic additivity of the effects of the two Lys residues was tested by comparing the behavior of the proteins with one or two internal Lys residues. Crystal structures were obtained for a few double Lys variants to examine structural consequences of burial of two internal Lys residues in the hydrophobic interior of the protein.

6.3 Materials and methods

6.3.1 Proteins

Variant proteins were created using the Stratagene Quickchange kit to introduce mutations into a hyperstable form of SNase known as Δ+PHS. Each protein was expressed in *E. coli* BL21/DE3 cells (Invitrogen) transformed with the plasmid Pet24A+. Proteins were expressed and purified by the method of Shortle and Meeker[69] as modified by García-Moreno et al[49].

6.3.2 Equilibrium thermodynamics

The Gibbs free energy of unfolding ($\Delta G°_{H2O}$) and pH_{mid} were measured using the intrinsic fluorescence of Trp-140 to monitor unfolding as described previously[71]. GdmCl (UltraPure grade, Invitrogen Life Technologies) was used as a denaturant. pH titrations were performed using 0.3 N HCl. Midpoints of pH-driven transitions were obtained by non-linear least squares fits with a two or three state model. All measurements were performed with an ATF-107 automated fluorometer (Aviv Inc.) at 25 ℃. The protein

concentration in these experiments was 50 µg/mL in a buffer consisting of 100mM NaCl with 25 mM CAPS, CHES, TAPS, or HEPES as appropriate.

6.3.3 CD spectroscopy

An Aviv model CD-420 circular dichroism spectrophotometer was used to collect far-UV CD spectra. Samples consisted of 50 µg/mL protein, 100 mM KCl and 25 mM TrisHCl (pH 7) or KAc (pH 5). Spectra were collected for a volume of 1 mL of protein sample in a 0.1 cm path-length quartz cuvette at 25 ℃. Measurements were taken at wavelength intervals of 1 nm using an averaging time of 3 s.

6.3.4 X-ray crystallography

The proteins were crystallized using the hanging drop vapor diffusion method. The reservoir solution contained 2-methyl-2,4-pentanedio (MPD) (Sigma-Aldrich Corp.) and 25 mM potassium phosphate. The protein was pre-incubated with 3'-5'-thymidine diphosphate (pdTp) and calcium chloride in a 1:2:3 molar ratio before mixing in a 1:1 ratio with the reservoir solution and incubated at 4 ℃. Crystals were flash cooled in liquid nitrogen. Diffraction data were collected using a Bruker Duo Apex diffractometer and processed using Bruker's software. Initial phasing for all structures was obtained by maximum likelihood-based molecular replacement method with Phaser software within the CCP4 suite using a previously solved structure for Δ+PHS (PDB ID: 3BDC) as a search model. Prior to molecular replacement, 3BDC.pdb was modified by truncating the substituted amino acid for the appropriate variant to Ala, removing all water molecules, and resetting all B-factors to 20.0 $Å^2$. Model building using Coot and refinement with

Refmac5 were performed iteratively to yield the final models. R-work and R-free residuals were monitored throughout the refinement. Water molecules were added during model building to reflect spherical electron density in 2Fo-Fc maps that were within 3.5 Å of a hydrogen bonding partner in the protein model. Final checks of the structures were done using SFCHECK and PROCHECK programs.

6.4 Results

6.4.1 Design principle and selection of Lys residues

Water is usually a better solvent for charges than the dry interior of a protein. Therefore, removal of a charge from water and burial in a protein is thermodynamically unfavorable. When an ionizable residue is buried in the interior of a protein, it can lead to a shift in the pK_a in the direction that promotes burial of the ionizable group in the neutral state[60]. For Lys residues, which in water titrate with pK_a values near 10, burial in the protein interior leads to the depression of its pK_a, in some cases by as many as 5 units[2]. This shift in pK_a renders the stability of the protein highly sensitive to pH. Starting at the pH corresponding to the normal pK_a of 10.4 for Lys in water, the stability of the folded protein decreases with decreasing pH at a rate of 1.36 kcal/mol (at 298 K). The stability decreases until a pH is reached where the Lys residue becomes charged.

A proof-of-principle study with a SNase variant with Lys-62 and Lys-125 showed that when two Lys residues are buried, the protein undergoes a highly cooperative transition between folded and unfolded states over a narrow range of pH near pH 7 (Chapter 5). When two Lys residues are buried in the same protein the stability of the protein decreases by 2 * 1.36 = 2.72 kcal/mol/pH. The pH where the conformational

transition is observed corresponds to the pH where $\Delta G°_{H2O} = 0$. This pH is determined by the relationship between the intrinsic, pH independent component of the stability of the protein, the magnitude of the shifts in pK_a, and the number of ionizable groups with shifted pK_a values. To achieve unfolding near pH 7, the decrease in stability originating with shifts in pK_a values has to be of magnitude comparable to the pH-independent component of stability.

The internal Lys residues previously studied were engineered into a highly stable variant of SNase[2], called Δ+PHS, which has a stability of 11.9 kcal/mol at pH 7. In the case of Lys-92, the ionization of the single Lys was sufficient to unfold SNase, but in most other variants the protein remained largely folded and native like under conditions of pH in which the internal Lys residues were charged. The effect of the ionizable residues on stability of the Lys-containing variant relative to the background protein ($\Delta\Delta G°_{H2O}$) can be described with:

$$\Delta\Delta G^{\circ}_{H2O} = \sum \Delta\Delta G^{\circ}_{H2O,mt} - \sum_i RT \ln \frac{1+e^{2.303\, z\, (\mathrm{pH} - pK_a^{D}{}_i)}}{1+e^{2.303\, z\, (\mathrm{pH} - pK_a^{N}{}_i)}} \qquad \text{Eq. 6.1}$$

where $\Delta\Delta G°_{H2O,mt}$ is the free energy difference between the reference protein and the variant with internal Lys under conditions of pH where the Lys is normally neutral (i.e. pH > 10). z is the charge of the internal ionizable side chain, and pK_a^{D} and pK_a^{N} are the pK_a values of the ionizable group in the denatured and native ensembles. The values of pK_a^{D} and pK_a^{N} are not coupled so this analysis does not suffer from uncertainties related to the properties in the denatured state. The rightmost term in Eq. 6.1 describes the contribution of ionizable residues to the pH-dependence of stability. This term shows that

the stability of the protein with a single Lys with anomalous pK_a changes by 1.36 kcal/mol for every pK_a unit difference between pK_a^D and pK_a^N.

The summation accounts for the fact that more than one group with anomalous pK_a can be present in the protein. Note that if two or three Lys with depressed pK_a values were present in the protein, its thermodynamic stability ($\Delta\Delta G°_{H2O}$) would decrease with decreasing pH with 2.72 or 4.08 kcal/mol/pH. Thus, by increasing the number of internal Lys residues with anomalous pK_a values the conformation of the protein can be made to shift from the folded to the unfolded state in the physiological range of pH, near 7. In the case of variants with buried Lys residues, the protein will be folded at high pH and unfolded at more acidic pH values. If Asp or Glu were buried, the trend would be opposite: folded in acidic pH and unfolded under more basic conditions.

The selection of the pairs of Lys residues for engineering the pH switches that would be active in the pH range 6 to 8 was based on their pK_a values and on the thermodynamic stability ($\Delta G°_{H2O}$) of the proteins with the buried Lys (Table 6.1). Simulations of the pH dependence of stability were performed using Eq. 6.1 based on the previously measured pK_a values and $\Delta G°_{H2O}$ and based on three assumptions: (1) the thermodynamic effects of the individual Lys residues are additive, (2) the pK_a value of the Lys residues in the double Lys variant remain the same as those measured in the single-Lys variants, and (3) the unfolding of the protein is cooperative between the fully folded and fully unfolded states. The simulations identified pairs of Lys residues that led to the condition of $\Delta G°_{H2O} = 0$ in the range pH 6 to 8. Twenty-five variants were selected to demonstrate that the concept behind the design of the original pH switch engineered with Lys-62 and Lys-125 was robust, that SNase could be converted into a pH switch

Table 6.1. Predicted and measured midpoints of pH unfolding of double Lys variants.

Variant	pK_a of 1st Lys[a]	pK_a of 2nd Lys[a]	pH_{mid} predicted	pH_{mid} corrected	1st pH_{mid}	2nd pH_{mid}[b]
V23K/T62K	7.3	8.1	7.9	7.5	4.3	-
L25K/T41K	6.3	9.3	6.8	5.1	5.2	4.7
L25K/T62K	6.3	8.1	8.0	7.5	7.7	4.6
L25K/A109K	6.3	9.2	8.3	7.5	7.3	-
L25K/L125K	6.3	6.2	9.0	8.7	8.7	-
L36K/T62K	7.2	8.1	7.5	7.1	5.5	-
L36K/I72K	7.2	8.6	8.8	8.6	8.7	-
L36K/V74K	7.2	7.4	8.7	8.7	8.5	-
L36K/L103K	7.2	8.2	8.4	8.2	6.8	-
L36K/A109K	7.2	9.2	7.9	6.8	6.4	-
T41K/V66K	9.3	5.6	6.5	6.6	5.2	-
T41K/I92K	9.3	5.3	7.8	7.0	6.5	-
T41K/V99K	9.3	6.5	7.8	6.8	6.2	-
T41K/L125K	9.3	6.2	6.7	5.1	5.7	-
T62K/V66K	8.1	5.6	7.8	7.8	5.7	4.6
T62K/V74K	8.1	7.4	8.4	7.3	7.9	6.5
T62K/A90K	8.1	8.6	8.4	7.6	8.3	6.1
T62K/Y91K	8.1	9.0	7.7	4.8	5.4	-
T62K/V104K	8.1	7.7	8.7	7.9	8.6	-
T62K/L125K	8.1	6.2	8.0	7.5	7.6	-
V66K/A109K	5.6	9.2	8.1	7.9	6.8	-
I72K/V74K	8.6	7.4	8.9	8.7	8.0	4.8
I72K/L103K	8.6	8.2	8.6	7.9	8.3	5.9
V74K/L103K	7.4	8.2	8.6	8.4	8.9	-
V74K/A109K	7.4	9.2	8.1	7.1	8.6	6.2

[a] Values from Isom et al. [2]

[b] More acidic pH_{mid} if data were fit using a three-state model

with many other pairs of internal Lys residues so long as their pK_a values were depressed.

6.4.2 Acid unfolding

The unfolding of the proteins as a function of pH measured with Trp fluorescence demonstrates that the majority of the double Lys variants of SNase act as pH switches near physiological values (Fig. 6.1, Table 6.1). Of the 25 variants examined, 17 unfolded in an apparent two-state manner; the other eight unfolded in an apparent three-state manner (Fig. 6.1). Some of the variants with apparent three-state unfolding were classified as such because the fluorescence data show two clear transitions (e.g. I72K/V74). Other variants were categorized as not being two-state based on the lack of an identifiable sharp transition between folded and unfolded baselines (e.g. L25K/T41K).

The pH_{mid} of unfolding was predicted for each protein using the previously measured pK_a of the Lys in the single Lys variants and known values for the stability of these variants at high pH where the Lys residue can be assumed to be neutral. Of the 17 proteins with apparent two-state behavior, 5 proteins unfold with pH_{mid} values within 0.5 pH units of the predicted values. Four of the 8 variants that unfolded in an apparent three-state manner also had a transition with pH_{mid} that was within 0.5 units of the expected value. Because the Δ+PHS background protein begins to unfold at pH > 10, the highest ΔG° measurements that were possible were done at pH values below the normal pK_a of 10.4 for Lys in water. When the predicted pH_{mid} values were corrected for the small error introduced by the inability to measure ΔG° at pH 11.4, an additional 5 of the two-state unfolding variants had pH_{mid} within 0.5 pH units of predicted values. Not only do the data (Fig. 6.1) confirm that most of the double Lys variants of SNase act as pH switches

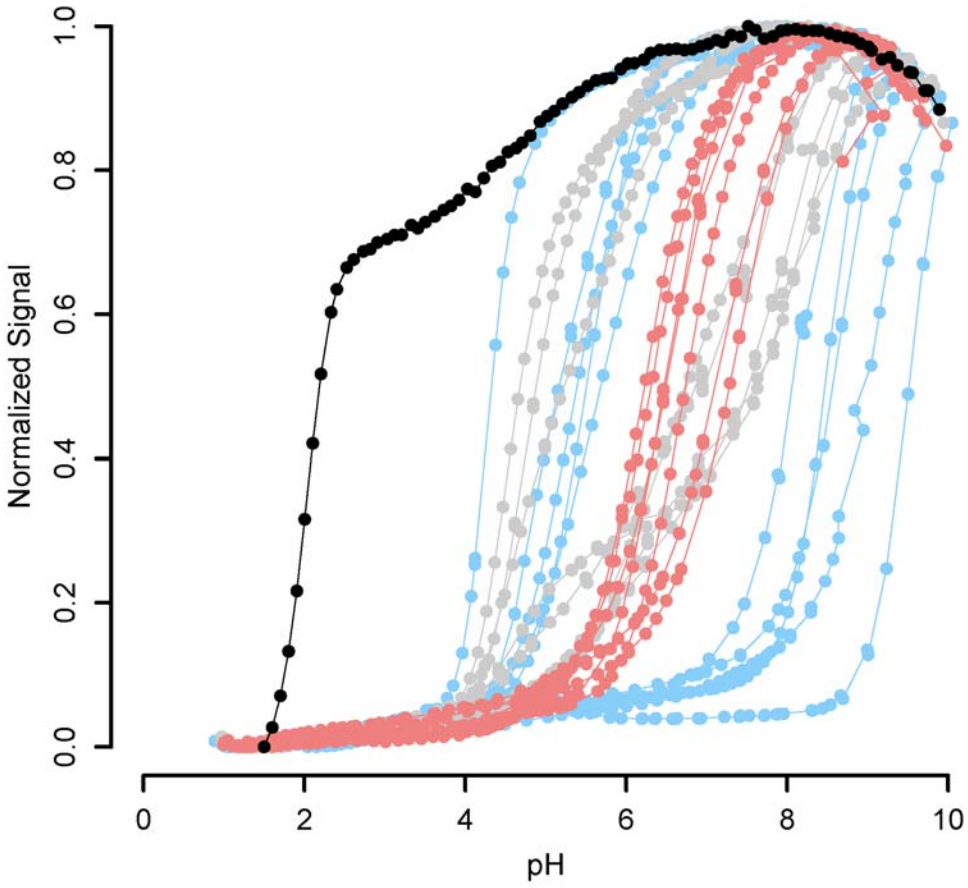

Figure 6.1 Acid titrations of variants of SNase with double Lys substitutions, monitored with Trp fluorescence. Acid unfolding of the background protein, Δ+PHS (black). Acid unfolding of double Lys variants with three-state unfolding (grey), two-state unfolding with pH_{mid} value between pH 6-8 (pink), and two-state unfolding with pH_{mid} values outside of pH 6-8 (blue). All traces are normalized to Δ+PHS.

that shift cooperatively between folded and unfolded states in response to small changes in pH in the physiological pH, the relatively good agreement between calculated and measured pH_{mid} values show that in most variants the properties of the individual Lys residues and their effect on the thermodynamic properties of the protein are additive.

6.4.3 pH dependence of thermodynamic stability

The pH-dependence of the difference in stability between variant and background protein ($\Delta\Delta G°_{H2O} = \Delta G°_{H2O,variant} - \Delta G°_{H2O,background}$) was measured (Table 6.2) because the slope of $\Delta\Delta G°_{H2O}$ versus pH is useful to distinguish between two possible mechanisms: (1) Both introduced Lys residues titrate with depressed pK_a values and the two residues ionize concomitant with global unfolding. In this situation the slope of $\Delta\Delta G°_{H2O}$ vs pH will decrease by 2.72 kcal/mol per pH unit. (2) One of the two Lys residues has a normal pK_a value because a subglobal conformational transition allows the Lys side chain to reach water. This destabilized and partially unfolded protein is then unfolded globally by the Lys with the depressed pK_a that gives rise to the pH-dependence of stability. In this the case the slope of $\Delta\Delta G°_{H2O}$ vs pH will decrease by 1.36 kcal/mol per pH unit.

Of the 25 proteins, 10 variants exhibited a pH-dependence consistent with two Lys residues with depressed pK_a; 9 variants had slopes consistent with only one Lys with depressed pK_a. The remaining 6 variants showed mixed behavior in which the $\Delta G°_{H2O}$ begins decreasing at high pH with a rate consistent with two Lys residue with depressed pK_a before the rate changes to a slope consistent with only one group with anomalous pK_a. It is possible that for these proteins, the two Lys residues have different pK_a values than for the single-site variants or that one of the Lys residues becomes solvated prior to

Table 6.2: Thermodynamic stability ($\Delta G°_{H2O}$) measured for the background protein (Δ+PHS) and Lys-containing variants of Δ+PHS.

Protein	pH	$\Delta G°_{H2O}$* (kcal/mol)	$\Delta\Delta G°_{H2O}$*,† (kcal/mol)
Δ+PHS	10.0	10.1 (0.1)	-
	9.9	10.4 (0.1)	-
	9.8	10.4 (0.1)	-
	9.5	11.3 (0.1)	-
	9.4	11.3 (0.1)	-
	9.1	11.5 (0.1)	-
	9.0	11.5 (0.1)	-
	8.9	11.6 (0.1)	-
	8.5	11.7 (0.1)	-
	8.0	11.9 (0.1)	-
	7.9	11.9 (0.1)	-
	7.5	11.8 (0.1)	-
	7.0	11.9 (0.1)	-
Δ+PHS/V23K/T62K	10.0	4.8 (0.2)	-5.6 (0.3)
	8.9	4.0 (0.1)	-7.6 (0.2)
	8.0	3.5 (0.1)	-8.4 (0.2)
Δ+PHS/L25K/T41K	10.0	4.8 (0.3)	-5.3 (0.4)
	9.0	4.2 (0.2)	-7.3 (0.3)
	8.1	3.2 (0.1)	-8.7 (0.2)
Δ+PHS/L25K/T62K	10.0	4.4 (0.1)	-5.7 (0.2)
	8.9	3.0 (0.1)	-8.6 (0.2)
	8.0	1.6 (0.1)	-10.3 (0.2)
Δ+PHS/L25K/A109K	10.0	2.9 (0.2)	-7.2 (0.3)
	8.9	2.1 (0.1)	-9.5 (0.2)
	8.0	1.2 (0.1)	-10.7 (0.2)
Δ+PHS/L25K/L125K	10.0	1.6 (0.2)	-8.5 (0.3)
	8.9	0.3 (0.1)	-11.3 (0.2)
Δ+PHS/L36K/T62K	9.9	4.8 (0.1)	-5.6 (0.2)
	8.9	4.4 (0.1)	-7.2 (0.2)
	8.0	3.5 (0.1)	-8.4 (0.2)
Δ+PHS/L36K/I72K	9.9	2.1 (0.7)	-8.3 (0.8)
	8.9	0.7 (0.2)	-10.9 (0.3)
Δ+PHS/L36K/V74K	10.0	2.2 (0.4)	-7.9 (0.5)
	8.9	0.4 (0.1)	-11.2 (0.2)
Δ+PHS/L36K/L103K	9.8	3.4 (0.1)	-7.0 (0.2)
	9.7	3.2 (0.2)	-7.6 (0.3)

	9.5	3.2	(0.1)	-8.1	(0.2)
	9.0	2.8	(0.1)	-8.7	(0.2)
	8.6	2.4	(0.2)	-9.3	(0.3)
	8.0	1.6	(0.5)	-10.3	(0.6)
Δ+PHS/L36K/A109K	10.0	3.8	(0.1)	-6.3	(0.2)
	9.5	3.6	(0.1)	-7.7	(0.2)
	9.0	3.4	(0.1)	-8.2	(0.2)
	8.6	3.0	(0.1)	-8.7	(0.2)
	8.0	2.6	(0.1)	-9.4	(0.2)
	7.5	1.8	(0.1)	-10.1	(0.2)
Δ+PHS/T41K/V66K	10.0	5.5	(0.1)	-4.6	(0.2)
	8.9	4.8	(0.1)	-6.8	(0.2)
	8.0	3.8	(0.1)	-8.1	(0.2)
Δ+PHS/T41K/I92K	9.8	3.0	(0.1)	-7.4	(0.2)
	9.4	3.2	(0.1)	-8.1	(0.2)
	9.0	3.3	(0.3)	-8.2	(0.4)
	8.6	2.8	(0.1)	-8.9	(0.2)
	8.0	2.0	(0.1)	-9.9	(0.2)
	7.5	1.1	(0.1)	-10.7	(0.2)
Δ+PHS/T41K/V99K	9.8	3.7	(0.1)	-6.7	(0.2)
	9.5	3.9	(0.1)	-7.5	(0.2)
	9.1	3.5	(0.1)	-8.0	(0.2)
	8.7	3.3	(0.1)	-8.3	(0.2)
	8.0	2.7	(0.1)	-9.2	(0.2)
	7.5	2.0	(0.1)	-9.8	(0.2)
Δ+PHS/T41K/L125K	9.9	4.8	(0.1)	-5.7	(0.2)
	9.8	4.8	(0.2)	-5.7	(0.3)
	9.0	4.5	(0.1)	-7.1	(0.2)
	8.1	3.2	(0.1)	-8.7	(0.2)
	7.1	1.9	(0.1)	-10.1	(0.2)
Δ+PHS/T62K/V66K	10.0	4.9	(0.1)	-5.2	(0.2)
	9.0	4.1	(0.1)	-7.4	(0.2)
	8.0	3.1	(0.1)	-8.8	(0.2)
Δ+PHS/T62K/V74K	9.9	3.9	(0.1)	-6.5	(0.2)
	8.9	2.6	(0.1)	-9.0	(0.2)
	8.0	0.9	(0.1)	-11.0	(0.2)
Δ+PHS/T62K/A90K	9.9	2.3	(0.1)	-8.1	(0.2)
	8.9	1.2	(0.1)	-10.4	(0.2)
Δ+PHS/T62K/Y91K	9.9	2.7	(0.1)	-7.7	(0.2)
	8.9	2.1	(0.2)	-9.5	(0.3)
	8.0	1.7	(0.2)	-10.2	(0.3)
	7.0	1.2	(0.2)	-10.7	(0.3)
Δ+PHS/T62K/V104K	9.8	2.8	(0.1)	-7.6	(0.2)

	8.9	1.5	(0.1)	-10.1	(0.2)
Δ+PHS/T62K/L125K	9.8	4.1	(0.1)	-6.3	(0.2)
	9.4	3.6	(0.1)	-7.6	(0.2)
	9.0	3.0	(0.1)	-8.5	(0.2)
	8.6	2.3	(0.1)	-9.4	(0.2)
	8.5	2.1	(0.1)	-9.6	(0.2)
	8.0	1.1	(0.2)	-10.8	(0.3)
Δ+PHS/V66K/A109K	9.7	3.4	(0.1)	-7.3	(0.2)
	9.6	3.4	(0.1)	-7.4	(0.2)
	9.0	3.0	(0.1)	-8.6	(0.2)
	8.6	2.5	(0.1)	-9.2	(0.2)
	8.0	2.1	(0.1)	-9.9	(0.2)
	7.5	1.3	(0.1)	-10.6	(0.2)
Δ+PHS/I72K/V74K	9.9	1.7	(0.2)	-8.7	(0.3)
	9.0	1.5	(0.2)	-10.1	(0.3)
Δ+PHS/I72K/L103K	9.9	2.0	(0.4)	-8.5	(0.5)
	9.0	1.1	(0.4)	-10.4	(0.5)
Δ+PHS/V74K/L103K	10.0	2.7	(0.5)	-7.4	(0.6)
	8.9	0.9	(0.2)	-10.7	(0.3)
Δ+PHS/V74K/A109K	9.9	2.8	(0.1)	-7.6	(0.2)
	8.9	2.0	(0.1)	-9.6	(0.2)
	8.0	1.0	(0.1)	-10.9	(0.2)
Δ+PHS/T41K/T62K/L103K	9.8	3.4	(0.1)	-7.0	(0.2)
	8.9	2.2	(0.1)	-9.4	(0.2)

* Fit errors are represented within parentheses

† $\Delta\Delta G°_{H2O} = \Delta G°_{H2O\ ,\ Variant} - \Delta G°_{H2O\ ,\ \Delta+PHS}$

the global unfolding of the protein at approximately the pH at which the rate of change in $\Delta G°_{H2O}$ occurs. It is also possible that limiting stability measurements to two pH values underestimates the number of proteins that exhibits this variable pH dependence.

Additivity in the properties of the two Lys residues can be examined more directly by comparing the sum of the effects of the individual Lys residues on the thermodynamic stability of the protein with the effects when the two Lys residues are incorporated into the same protein. In 13 of the 25 variants, the experimental $\Delta G°_{H2O}$ at pH 10 was within error of the $\Delta G°_{H2O}$ predicted by combining the value measured for the proteins with the single Lys residues. For 12 of the proteins, the measured stabilities were either higher or lower than the expected values, suggesting that in half of the variants the Lys residues do not behave with additivity. However, in no case are the deviations between expected and measured values large. The relatively high additivity suggests that in most variants the native fold is retained at least at high pH where the Lys residues are not expected to be charged and the effects from shifts in pK_a are expected to be small.

6.4.4 Structural studies with CD spectroscopy

To determine the extent to which the Lys substitutions affect the global structure of the protein, far-UV CD spectra were collected for 14 of the variants at pH values above and below the respective pH_{mid} values. With the exception of the variant with T62K/Y91K, which is known to give anomalous far UV-CD spectra[36,60,73] (Chapters 2 and 3), the CD data of all of the variants collected near pH 9, well above the measured pH_{mid} values, show that the proteins are fully folded and comparable to the Δ+PHS protein (Fig. 6.2A). The CD spectra collected near or below the pH_{mid} suggest that under

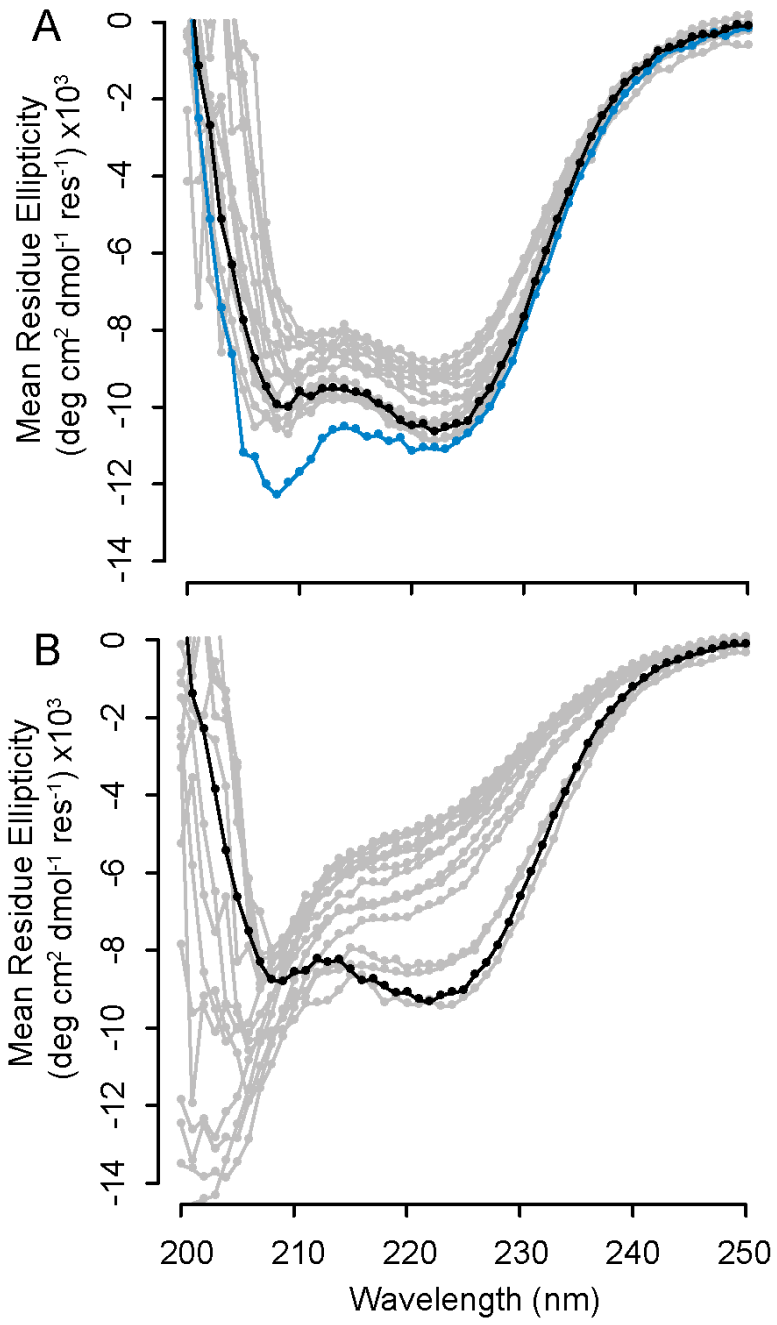

Figure 6.2 Far-UV CD spectra for double lysine variants. The background protein (Δ+PHS) is shown in black, double lysine variants are shown in grey and the variant with T62K/Y91K is shown in blue. **(A)** high pH conditions near pH 9, and **(B)** low pH conditions at or below the pH_{mid} of the protein variants.

these more acidic conditions the protein is indeed fully unfolded or as unfolded as SNase becomes in water (Fig. 6.2B). The CD spectra demonstrate that the transition triggered by the change in pH is between two very different conformational states.

6.4.5 Crystal structures

Crystal structures were solved primarily to demonstrate that burial of two Lys residues in the neutral state does not change the conformation of the protein in a detectable manner and to examine the microenvironments of the buried Lys side chains and the potential for interaction between them. Structures were determined for variants with I72K/V74K, V66K/A109K, and L36K/L103K substitutions (Fig. 6.3, Fig. 6.4, Table A5).

The structure of the variant with I72K/V74K was solved at both pH 7 and pH 9, bracketing one of the acid unfolding transitions, to 1.65 Å and 1.60 Å, respectively. In both structures, there is a slight perturbation to the β-barrel manifested as a shift of β-strands 1-3 away from α-helix-1. The tertiary structure is comparable to that of the Δ+PHS protein used as background for the Lys substitutions (PDB: 3BDC) with an overall C_α RMSD of 0.3 Å relative to 3BDC. For the structure solved at pH 7, there are water molecules approximately 3 Å from the amino groups of Lys-72 and Lys-74. Both Lys-72 and Lys-74 are also 2.8 Å away from the carbonyl oxygen atoms of Ala-69 and Pro-11, respectively. The solvent accessible surface area (SASA) for the Nζ of Lys-72 is 24.1 $Å^2$ (52%) and 6.3 $Å^2$ (12%) for Lys-74. The structure solved at pH 9 is very similar to the pH 7 structure. The Nζ atom of Lys-74 has one nearby water 2.8 Å away and is 3.0 Å away from the carbonyl oxygen on Pro-11. The Nζ atom of Lys-72 points towards

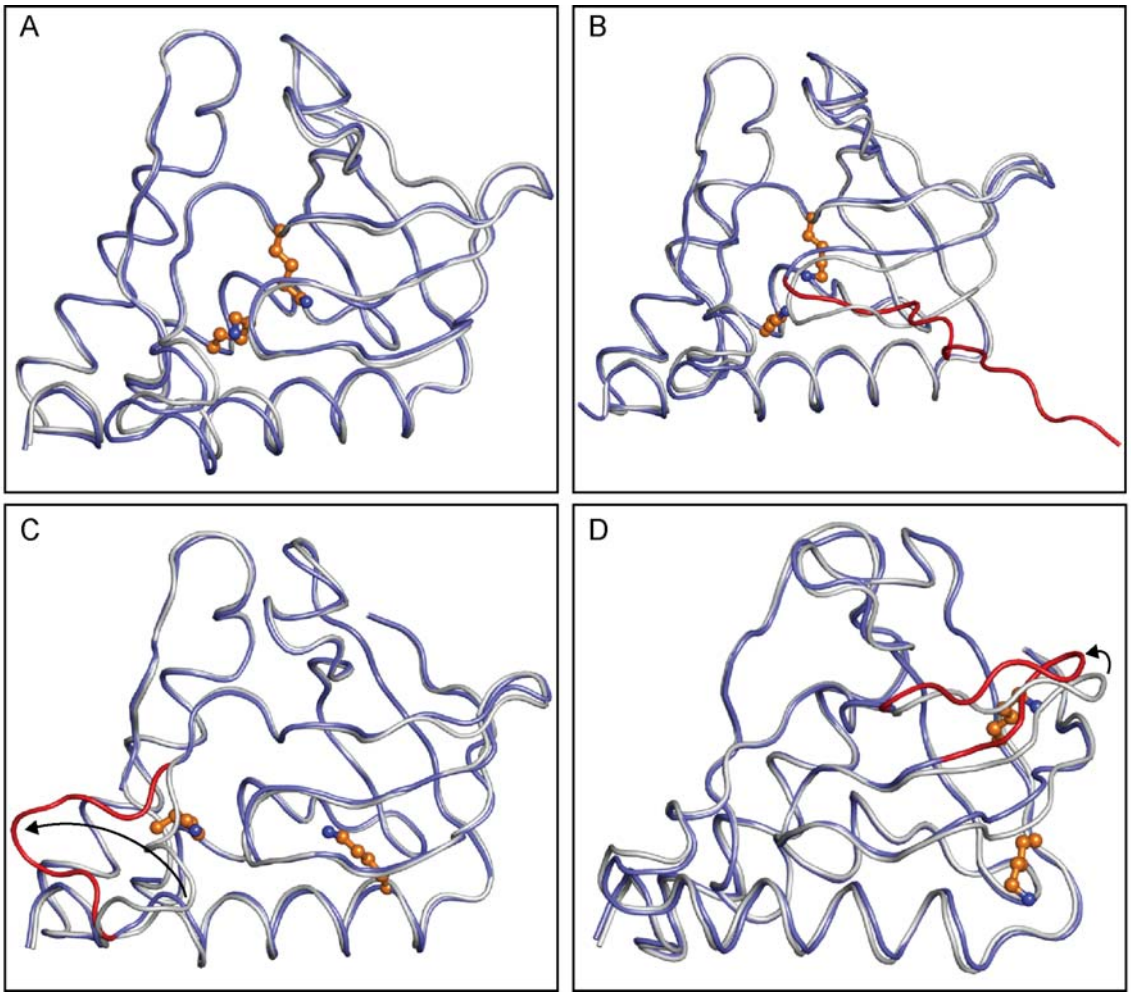

Figure 6.3 Crystal structures of double Lys variants. Overlay of backbone traces of variants (blue) and Δ+PHS nuclease (white). **(A)** L36K/L103K (monomer) (PDB ID 5I9O), **(B)** L36K/L103K (monomer from a domain swapped variant), **(C)** V66K/A109K, **(D)** I72K/V74K (pH 7) (PDB ID 4PMC). Notable structural differences are highlighted in red. Lys residues of interest are shown as ball and stick in orange.

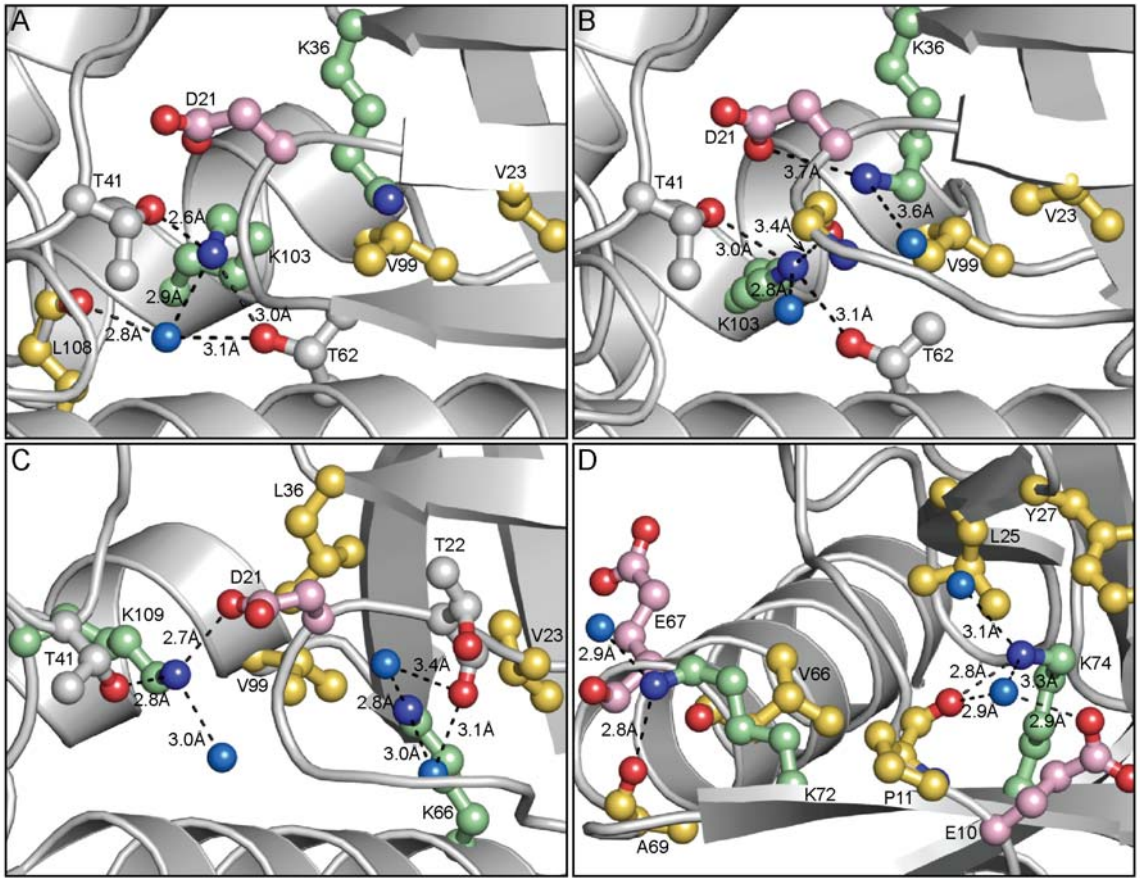

Figure 6.4 Microenvironments of substituted Lys residues. (A) Δ+PHS/L36K/L103K (pH 7) **(B)** Δ+PHS/L36K/L103K (pH 6), **(C)** Δ+PHS/V66K/A109K, and **(D)** Δ+PHS/I72K/V74K. Nearby residues are represented as ball and stick with Lys shown in green, hydrophobic residues in yellow, and acidic residues shown in pink. Water molecules are depicted as light blue spheres. For clarity, β-strand 1 in D) and β-strands 1 and 2 in C) are shown as loops.

the end of α-helix 1, which is different from what is observed in the pH 7 structure. There is one water molecule 2.9 Å from the amino of Lys-72, and three carbonyl oxygen atoms (Val-66, Glu-67, and Ala-69) within 3.5 Å of the amine. In this high pH structure, solvent accessible surface area for the Nζ of Lys-72 is 6.8 $Å^2$ (13%) and 3.7 $Å^2$ (7%) for Lys-74.

In the 1.45 Å structure of the variant with V66K/A109K, the truncated omega loop consisting of residues 40-52 is flipped towards α-helix 3 relative to the orientation observed in the Δ+PHS protein used as background. The overall C_α RMSD is 0.3 Å relative to 3BDC. The Nζ of Lys-66 has 0.4 $Å^2$ (0.7%) of solvent accessibility while Lys-109 has 3.7 $Å^2$ (6%) accessible so these are both buried Lys residues. Each internal Lys also has a single water molecule 3.0 Å from the amine. The Nζ of Lys-109 is 2.7 Å away from the Oδ1 of Asp-21 and 2.8 Å from the hydroxyl group on Thr-41.

Two structures were solved for the variant with the L36K/L103K substitutions. At pH 7, the protein is monomeric with two copies in the asymmetric unit. At pH 6, below the pH_{mid} of this protein, the protein forms a homodimer in the asymmetric unit. The structure at pH 7, refined to 1.95 Å, has no significant differences between the two monomers within the asymmetric unit (C_α RMSD = 0.1). The Nζ of Lys-36 has a SASA of 1.1 $Å^2$ (2.1%) and the Nζ of Lys-103 has 0.1 $Å^2$ (0.2%) exposed to bulk solvent. There are no hydrogen bonding partners near Lys-36. Lys-103 has three potential interactions with a water molecule 2.6 Å away and the hydroxyl moieties on Thr-41 (2.6 Å) and Thr-62 (3.0 Å). The 2.05 Å structure of Δ+PHS/L36K/L103K at pH 6 forms a homodimer in the asymmetric unit wherein residues 7-17 domain swap to form the majority of β-strand in the adjacent monomer. No significant differences were observed between the monomers (C_α RMSD = 0.3). Omitting the domain swapped portion, the overall tertiary

structure of the rest of the protein is unchanged with a C_α RMSD of 0.4 Å relative to 3BDC. The Nζ of both Lys-36 and Lys-103 are internal with less than 1% surface area accessible to bulk solvent. Lys-103 points towards the carbonyl of Gly-20, 3.4 Å away, and the Nζ atom is 3.1 Å and 3.0 Å away from the hydroxyl of Thr-62 and Thr-41, respectively. There is a water molecule 2.5 Å away from the Nζ of Lys-103. There is nothing within 3.5 Å of the Nζ atom of Lys-36, but there is a water molecule 3.6 Å away not seen in other structures containing the same mutation.

6.5 Discussion

This study demonstrates that internal Lys residues with depressed pK_a values can be used to engineer pH switch proteins that can undergo a large conformational transition in response to a small change in pH in the physiological range. The fact that SNase can be turned into a pH switch with many different pairs of Lys residues demonstrates this is a robust approach. From the knowledge of the pK_a values of internal residues, it is even possible to predictably engineer cooperative pH switching behavior from a fully folded to fully unfolded states.

Of the 25 proteins created to test this approach, 6 variants (25/125, 36/72, 36/74, 62/104, 62/125, and 74/103) met all three criteria of (1) having a pH_{mid} within error of predicted value based on the behavior of the individual Lys substitutions, (2) undergoing cooperative unfolding, and (3) global unfolding was driven by the titration of both introduced Lys residues. Half of the variants tested, including those designed specifically to challenge the design principles, had a pH_{mid} close to the predicted value, emphasizing that even when the unfolding was not two-state or driven by only one Lys residues, the

thermodynamic approach used in this study can indeed be used to select mutation sites for engineering pH sensitive behavior. It is noteworthy that in some cases the design did not work as intended. These cases will be studied further to obtain insight into the limitations of this approach and potential problems when applied to other proteins.

6.5.1 Interactions between sites

When the sites for substitution with Lys were selected, it was assumed that the pK_a values of the Lys residues remained the same in the double variant as in the single variants and that the contributions of the individual Lys resides would be purely additive. Both of these assumptions should fail if the two Lys are situated close together in the tertiary structure - the charged state of one Lys would affect the other. To determine the extent to which the distance between the two Lys plays into the model, variants with mutations at different Lys-Lys distances were compared. Variants such as T62K/V66K, in which the Lys are close in both sequence and space, and variants such as V23K/T62K, where the Lys side chains are only near in space, are expected to deviate from predictions if the titration of the internal Lys are coupled. Possible indicators of such coupled titrations include a pH_{mid} significantly different than predicted, multi-state unfolding, unfolding driven by the titration of only one Lys residue, and non-additivity in thermodynamic stability.

The distance between the amino groups of each Lys residue in the double Lys variants was predicted based on available crystal structures of single Lys variants. In general, the double Lys variants with a predicted distance between Lys amino groups > 10 Å met the three design criteria while variants with Lys-Lys distances predicted to be

< 10 Å failed to meet at least one criterion. In these proteins, either the acid unfolds the protein a non two-state manner, the unfolding happens at a pH_{mid} far from the predicted one, or there is evidence that only one of the two groups binds H^+ coincident with the acid unfolding transition. Some variants, for example I72K/L103K and T62K/V74K with predicted distances between Lys amino groups of 16.5 Å and 12.9 Å, respectively, also exhibited apparent three-state unfolding during acid titration experiments. Interestingly, the pK_a values measured for the Lys residues individually are near the more basic of the two pH_{mid} values measured. This may indicate that in the case of these proteins both Lys residues become protonated prior to the global unfolding of the protein. It is possible that the titration of the Lys residues led to a significant yet incomplete shift towards an unfolded state, stopping at a stable yet partially unfolded state.

6.5.2 Cooperativity

The proof of principle experiment with Lys-62 and Lys-125 demonstrating that a variant of SNase with two internal Lys residues with anomalous pK_a values could be unfolded cooperatively was surprising (Chapter 5). Previous studies of SNase variants with individual Lys residues have shown that the Lys side chains are tolerated in the protein interior without the need for any structural adaptation, but that as the pH decreases below the pH corresponding to the normal pK_a of 10.4 of Lys in water, conformational reorganization can ensue[52,53,108,123]. It was not obvious a priori that the substitution of a second internal position with Lys would not destroy the cooperative core of the protein. The variant with two internal Lys residues could well have responded with a slight conformational change that would lead to the normalization of one of the two

anomalous pK_a values, thereby eliminating the role of one of the two Lys residues as pH sensor, and lead to an unstable and partially unfolded protein that would unfold without cooperativity. But as the data with 25 other double Lys variants of SNase show, that is not what happened. In most variants, the protein, even after substitution of two internal positions with Lys, was still native-like at pH 9 and capable of undergoing a cooperative transition. This is consistent with the observation that by and large, the effects of each of the two internal Lys residues are additive. Regardless of the how the Lys side chains might interact with each other, most still behave as anomalous ionizable residues indicating that they are still buried in the protein interior and capable of driving the acid unfolding transition.

In this engineering exercise, it was assumed that partially unfolded intermediates would not be populated significantly during the acid unfolding reaction because partially unfolded states are usually higher in energy than the fully unfolded state. This ignored the fact that in variants of SNase with the internal Val-23 substituted with Glu or Lys populate a partially unfolded state in which the β1-β2 strands are flipped out[53,58]. Any substitution that decreases the energy gap between the folded and unfolded states can potentially alter the entire landscape of accessible partially folded states [115,163] and the substitutions with Lys could have lead to population of a partially unfolded state. While the crystal structures do suggest that two internal Lys residues lead to slight structural reorganization of the protein, the protein still unfolds cooperatively despite the fact that a small number of H^+ are bound upon unfolding.

When ionizable groups are introduced to positions 23, 66, and 109, the protein undergoes structural reorganization concomitant with ionization of these

groups[1,52,58,108,123]. The conformational reorganization allows the ionizable groups at these positions to contact water. These positions were chosen to act as control positions and were used in the creation of six variants. All six of these variants fell into the group of proteins in which only one of the two Lys residues contribute actively to the acid unfolding behavior by binding of an H^+. Interestingly, all four proteins containing the T41K mutation also behaved in this manner. It is highly probable that a mutation at position 41 causes the truncated omega loop in the protein to swing in a similar fashion as is seen with mutations at position 109 thus exposing the nearby interior of the protein to water. The crystal structure of the variant with V66K/A109K illustrates this specific structural response as both it and the structure of a variant with A109R show identical positioning of the truncated omega loop and subsequent invasion of the protein interior with water molecules[1].

6.6 Conclusions

The engineering of pH switch proteins based on use of buried Lys residues with anomalous pK_a values could have failed for a number of reasons. Yet, remarkably, the 25 pairs of internal Lys that were selected led to behavior consistent with that of a pH switch that drives a large conformational transition in response to a small change in pH in the physiological range. In some cases the two Lys residues bound H^+ concomitant with the unfolding reaction, but in some cases a pre-unfolding transition leads to the normalization of the pK_a of one of the two Lys residues. All things considered, the behavior of the two Lys residues was highly additive, suggesting that important properties of the pH switches,

such as the range of pH where the switch can be expected to be active, can be predicted based on known properties of the individual Lys residues.

The experimental data demonstrate that the design principle based on buried Lys residues with anomalous pK_a values is robust. It is also general and transferable to other proteins, for two reasons. First, because the design takes advantage of the relatively dry character of the protein interior, which is common to all proteins and essential for the depression in pK_a of Lys residues required to achieve pH sensing capabilities at physiological pH values. Second, because the design is based on the principles of linkage thermodynamics, which govern the coupling between equilibrium processes involving H^+ binding and conformational transitions. All expectations are that this approach can be used to turn any protein into a pH sensing protein active in the physiological range of pH.

References

1. Harms, M. J., Schlessman, J. L., Sue, G. R. & E, B. G. Arginine residues at internal positions in a protein are always charged. *PNAS* **108,** 18954–18959 (2011).
2. Isom, D. G., Castañeda, C. A., Cannon, B. R. & García-Moreno E., B. Large shifts in pKa values of lysine residues buried inside a protein. *PNAS* **108,** 5260–5265 (2011).
3. Thanki, N., Thornton, J. M. & Goodfellow, J. M. Influence of secondary structure on the hydration of serine, threonine and tyrosine residues in proteins. *Protein Eng.* **3,** 495–508 (1990).
4. Nick Pace, C., Martin Scholtz, J. & Grimsley, G. R. Forces stabilizing proteins. *FEBS Lett.* **588,** 2177–2184 (2014).
5. Loladze, V. V., Ermolenko, D. N. & Makhatadze, G. I. Thermodynamic consequences of burial of polar and non-polar amino acid residues in the protein interior. *J. Mol. Biol.* **320,** 343–357 (2002).
6. Takano, K., Yamagata, Y. & Yutani, K. Contribution of polar groups in the interior of a protein to the conformational study. *Biochemistry* **40,** 4853–4858 (2001).
7. Bartlett, G. J., Porter, C. T., Borkakoti, N. & Thornton, J. M. Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.* **324,** 105–121 (2002).
8. Ekici, O. D., Paetzel, M. & Dalbey, R. E. Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci.* **17,** 2023–2037 (2008).
9. Decker, T. & Kovarik, P. Serine phosphorylation of STATs. *Oncogene* **19,** 2628–

2637 (2000).

10. Shi, Y. Serine/threonine phosphatases: mechanism through structure. *Cell* **139,** 468–84 (2009).

11. Kisselev, A. F., Songyang, Z. & Goldberg, A. L. Why does threonine, and not serine, function as the active site nucleophile in proteasomes? *J. Biol. Chem.* **275,** 14831–14837 (2000).

12. Schönichen, A., Webb, B. a, Jacobson, M. P. & Barber, D. L. Considering protonation as a posttranslational modification regulating protein structure and function. *Annu. Rev. Biophys.* **42,** 289–314 (2013).

13. Li, P., Martins, I. R. S., Amarasinghe, G. K. & Rosen, M. K. Internal dynamics control activation and activity of the autoinhibited Vav DH domain. *Nat. Struct. Mol. Biol.* **15,** 613–618 (2008).

14. Goncearenco, A. & Berezovsky, I. N. Protein function from its emergence to diversity in contemporary proteins. *Phys. Biol.* **12,** 045002 (2015).

15. Gutteridge, A. & Thornton, J. M. Understanding nature's catalytic toolkit. *Trends Biochem. Sci.* **30,** 622–629 (2005).

16. Shen, A. L., Sem, D. S. & Kasper, C. B. Mechanistic studies on the reductive half-reaction of NADPH-cytochrome P450 oxidoreductase. *J. Biol. Chem.* **274,** 5391–5398 (1999).

17. Hubbard, P. a., Shen, A. L., Paschke, R., Kasper, C. B. & Kim, J. J. P. NADPH-cytochrome P450 oxidoreductase. Structural basis for hydride and electron transfer. *J. Biol. Chem.* **276,** 29163–29170 (2001).

18. Rastogi, V. K. & Girvin, M. E. Structural changes linked to proton translocation

by subunit c of the ATP synthase. *Nature* **402,** 263–268 (1999).

19. Aksimentiev, A., Balabin, I. A., Fillingame, R. H. & Schulten, K. Insights into the molecular mechanism of rotation in the Fo sector of ATP synthase. *Biophys. J.* **86,** 1332–1344 (2004).

20. Pisliakov, A. V, Sharma, P. K., Chu, Z. T., Haranczyk, M. & Warshel, A. Electrostatic basis for the unidirectionality of the primary proton transfer in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7726–7731 (2008).

21. Koepke, J. *et al.* High resolution crystal structure of Paracoccus denitrificans cytochrome c oxidase: new insights into the active site and the proton transfer pathways. *Biochim. Biophys. Acta* **1787,** 635–645 (2009).

22. Chatake, T. *et al.* Protonation states of buried histidine residues in human deoxyhemoglobin revealed by neutron crystallography. *J. Am. Chem. Soc.* **129,** 14840–14841 (2007).

23. Bhattacharya, S., Sukits, S. F., MacLaughlin, K. L. & Lecomte, J. T. The tautomeric state of histidines in myoglobin. *Biophys. J.* **73,** 3230–40 (1997).

24. Klug, A. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q. Rev. Biophys.* **43,** 1–21 (2010).

25. Liao, S.-M., Du, Q.-S., Meng, J.-Z., Pang, Z.-W. & Huang, R.-B. The multiple roles of histidine in protein interactions. *Chem. Cent. J.* **7,** 44 (2013).

26. Zheng, G., Schaefer, M. & Karplus, M. Hemoglobin Bohr effects: Atomic origin of the histidine residue contributions. *Biochemistry* (2013). doi:10.1021/bi401126z

27. Cerdà-Costa, N. & Gomis-Rüth, F. X. Architecture and function of

metallopeptidase catalytic domains. *Protein Sci.* **23,** 123–144 (2014).

28. Perutz, M. F. Stereochemistry of cooperative effects in haemoglobin. *Nature* **228,** 726–739 (1970).

29. Mattevi, A. *et al.* Crystal structure of Escherichia coli pyruvate kinase type I: molecular basis of the allosteric transition. *Structure* **3,** 729–741 (1995).

30. Waldburger, C. D., Schildbach, J. F. & Sauer, R. T. Are buried salt bridges important for protein stability and conformational specificity? *Nat. Struct. Biol.* **2,** 122–128 (1995).

31. Takeuchi, H., Okada, A. & Miura, T. Roles of the histidine and tryptophan side chains in the M2 proton channel from influenza A virus. *FEBS Lett.* **552,** 35–38 (2003).

32. Maity, H., Lim, W. K., Rumbley, J. N. & Englander, S. W. Protein hydrogen exchange mechanism: local fluctuations. *Protein Sci.* **12,** 153–160 (2003).

33. Thurlkill, R. L., Grimsley, G. R., Scholtz, J. M. & Pace, C. N. pK values of the ionizable groups of proteins. *Protein Sci* **15,** 1214–1218 (2006).

34. Castañeda, C. A. *et al.* Molecular determinants of the pKa values of Asp and Glu residues in staphylococcal nuclease. *Proteins* **77,** 570–88 (2009).

35. Isom, D. G., Castañeda, C. A., Cannon, B. R., Velu, P. D. & García-Moreno E., B. Charges in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 16096–100 (2010).

36. Cannon, B. R. Thermodynamic consequences of substitutions of internal positions in proteins with polar and ionizable residues. (2008).

37. Fujiwara, K., Toda, H. & Ikeguchi, M. Dependence of α-helical and β-sheet amino

acid propensities on the overall protein fold type. *BMC Struct. Biol.* **12,** 18 (2012).

38. Richman, D. E. Conformational responses to changes in the state of ionization of titratable groups in proteins. (2015).

39. Shi, C., Wallace, J. a & Shen, J. K. Thermodynamic coupling of protonation and conformational equilibria in proteins: theory and simulation. *Biophys. J.* **102,** 1590–7 (2012).

40. Fitch, C. A., Platzer, G., Okon, M., García-Moreno E., B. & McIntosh, L. P. Arginine: Its pKa value revisited. *Protein Sci.* **24,** 752–61 (2015).

41. Grimsley, G. R., Scholtz, J. M. & Pace, C. N. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci.* **18,** 247–251 (2009).

42. Hansen, A. L. & Kay, L. E. Measurement of histidine pKa values and tautomer populations in invisible protein states. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E1705–12 (2014).

43. Tanokura, M. 1H-NMR study on the tautomerism of the imidazole ring of histidine residues. II. Microenvironments of histidine-12 and histidine-119 of bovine pancreatic ribonuclease A. *Biochim. Biophys. Acta* **742,** 586–596 (1983).

44. Stickle, D. F., Presta, L. G., Dill, K. A. & Rose, G. D. Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226,** 1143–1159 (1992).

45. Tucker, P. W., Hazen, E. E. & Cotton, F. A. Staphylococcal nuclease reviewed: a prototypic study in contemporary enzymology. II. Solution studies of the nucleotide binding site and the effects of nucleotide binding. *Mol. Cell. Biochem.* **23,** 3–16 (1979).

46. Tucker, P. W., Hazen, E. E. & Cotton, F. A. Staphylococcal nuclease reviewed: a

prototypic study in contemporary enzymology. IV. The nuclease as a model for protein folding. *Mol. Cell. Biochem.* **23,** 131–141 (1979).

47. Tucker, P. W., Hazen, E. E. & Cotton, F. A. Staphylococcal nuclease reviewed: a prototypic study in contemporary enzymology. I. Isolation; physical and enzymatic properties. *Mol. Cell. Biochem.* **22,** 67–77 (1978).

48. Tucker, P. W., Hazen, E. E. & Cotton, F. A. Staphylococcal nuclease reviewed: a prototypic study in contemporary enzymology. III. Correlation of the Three-Dimensional Structure with the Mechanisms of Enzymatic Action. *Mol. Cell. Biochem.* **23,** 67–86 (1979).

49. García-Moreno E., B. *et al.* Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* **64,** 211–24 (1997).

50. Baran, K. L. *et al.* Electrostatic effects in a network of polar and ionizable groups in staphylococcal nuclease. *J. Mol. Biol.* **379,** 1045–62 (2008).

51. Doctrow, B. ELECTROSTATIC COUPLING AND CONFORMATIONAL FLUCTUATIONS AS DETERMINANTS OF PKA VALUES IN PROTEINS by. (2014).

52. Chimenti, M. S. *et al.* Structural reorganization triggered by charging of Lys residues in the hydrophobic interior of a protein. *Structure* **20,** 1071–1085 (2012).

53. Richman, D. E., Majumdar, A. & E, B. G. pH Dependence of Conformational Fluctuations of the Protein Backbone. *Proteins Struct. Funct. Bioinforma.* 3132–3143 (2014). doi:10.1002/prot.

54. Khangulov, V. S. Molecular determinants of pKa values of internal ionizable groups in proteins: Interactions of internal ionizable groups with surface charges.

(2011).

55. Harms, M. J. *et al.* The pK(a) values of acidic and basic residues buried at the same internal location in a protein are governed by different factors. *J. Mol. Biol.* **389,** 34–47 (2009).

56. Baker, E. N. & Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **44,** 97–179 (1984).

57. Thanki, N., Thornton, J. M. & Goodfellow, J. M. Distributions of water around amino acid residues in proteins. *J. Mol. Biol.* **202,** 637–657 (1988).

58. Robinson, A. C. Molecular determinants of dielectric properties of proteins examined with internal ionizable residues and buried ion pairs. (2012).

59. Murtaugh, M. L., Fanning, S. W., Sharma, T. M., Terry, A. M. & Horn, J. R. A combinatorial histidine scanning library approach to engineer highly pH-dependent protein switches. *Protein Sci.* **20,** 1619–31 (2011).

60. Isom, D. G., Cannon, B. R., Castañeda, C. A., Robinson, A. C. & García-Moreno E., B. High tolerance for ionizable residues in the hydrophobic interior of proteins. *PNAS* **105,** 17784–8 (2008).

61. Isom, D. G., Castañeda, C. A., Cannon, B. R., Velu, P. D. & García-Moreno E., B. Charges in the hydrophobic interior of proteins. *PNAS* **107,** 16096–100 (2010).

62. Jiang, Y., Ruta, V., Chen, J., Lee, A. & MacKinnon, R. The principle of gating charge movement in a voltage-dependent K+ channel. *Nature* **423,** 42–48 (2003).

63. Gloeckner, C. J. *et al.* The Parkinson disease causing LRRK2 mutation I2020T is associated with increased kinase activity. *Hum. Mol. Genet.* **15,** 223–232 (2006).

64. Lahiry, P. *et al.* A mutation in the serine protease TMPRSS4 in a novel pediatric

neurodegenerative disorder. *Orphanet J. Rare Dis.* **8,** 126 (2013).

65. Xie, T. X. *et al.* Serine substitution of proline at codon 151 of TP53 confers gain of function activity leading to anoikis resistance and tumor progression of head and neck cancer cells. *Laryngoscope* **123,** 1416–1423 (2013).

66. Ayuso-Tejedor, S., Abián, O. & Sancho, J. Underexposed polar residues and protein stabilization. *Protein Eng. Des. Sel.* **24,** 171–177 (2011).

67. Blaber, M., Lindstrom, J. & Gassner, N. Energetic Cost and Structural Consequences of Burying a Hydroxyl Group within the Core of a Protein Determined from Ala-> Ser and Val->Thr Substitutions in T4 Lysozyme. *Biochemistry* **32,** 11363–11373 (1993).

68. Takano, K., Scholtz, J. M., Sacchettini, J. C. & Pace, C. N. The Contribution of Polar Group Burial to Protein Stability Is Strongly Context-dependent. *J. Biol. Chem.* **278,** 31790–31795 (2003).

69. Shortle, D. & Meeker, A. K. Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. *Proteins* **1,** 81–9 (1986).

70. Byrne, M. P., Manuel, R., Lowe, L. & Stites, W. E. Energetic Contribution of Side Chain Hydrogen Bonding to the Stability of Staphylococcal Nuclease. *Biochemistry* **34,** 13949–13960 (1995).

71. Whitten, S. T. & García-Moreno E., B. pH Dependence of Stability of Staphylococcal Nuclease: Evidence of Substantial Electrostatic Interactions in the Denatured State †. *Biochemistry* **39,** 14292–14304 (2000).

72. Karp, D. A. Structural and energetic consequences of the ionization of internal groups in staphylococcal nuclease. (2007).

73. Isom, D. G. pKa values of internal ionizable groups in staphylococcal nuclease. (2006).

74. Malkov, S. N., Živković, M. V., Beljanski, M. V., Hall, M. B. & Zarić, S. D. A reexamination of the propensities of amino acids towards a particular secondary structure: Classification of amino acids based on their chemical structure. *J. Mol. Model.* **14,** 769–775 (2008).

75. Gromiha, M. M., Oobatake, M., Kono, H., Uedaira, H. & Sarai, a. Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations. *Protein Eng.* **12,** 549–55 (1999).

76. Shortle, D., Stites, W. E. & Meeker, A. K. Contributions of the Large Hydrophobic Amino Acids to the Stability of Staphylococcal Nuclease. *Biochemistry* **29,** 8033–8041 (1990).

77. Shortle, D. Staphylococcal nuclease: A showcase of m-value effects. *Adv. Protein Chem.* **46,** 217–249 (1995).

78. Harms, M. J. Molecular determinants of pKa values of internal ionizable groups in Staphylococcal nuclease. (2008).

79. Choi, K. H. *et al.* New superfamily members identified for Schiff-Base enzymes based on verification of catalytically essential residues. *Biochemistry* **45,** 8546–8555 (2006).

80. Luecke, H. Atomic resolution structures of bacteriorhodopsin photocycle intermediates: The role of discrete water molecules in the function of this light-driven ion pump. *Biochim. Biophys. Acta - Bioenerg.* **1460,** 133–156 (2000).

81. Zscherp, C., Schlesinger, R., Tittor, J., Oesterhelt, D. & Heberle, J. In situ determination of transient pKa changes of internal amino acids of bacteriorhodopsin by using time-resolved attenuated total reflection Fourier-transform infrared spectroscopy. *PNAS* **96,** 5498–503 (1999).

82. Yoshikawa, S., Muramoto, K. & Shinzawa-Itoh, K. Proton-pumping mechanism of cytochrome C oxidase. *Annu. Rev. Biophys.* **40,** 205–223 (2011).

83. von Ballmoos, C., Wiedenmann, A. & Dimroth, P. Essentials for ATP synthesis by F1F0 ATP synthases. *Annu. Rev. Biochem.* **78,** 649–672 (2009).

84. Ge, X. & Gunner, M. R. Unraveling the mechanism of proton translocation in the extracellular half-channel of bacteriorhodopsin. *bioRxiv* **1,** 639–654 (2015).

85. Vorburger, T. *et al.* Arginine-induced conformational change in the c-ring/a-subunit interface of ATP synthase. *FEBS J.* **275,** 2137–2150 (2008).

86. Plesniak, L. a, Connelly, G. P., Wakarchuk, W. W. & McIntosh, L. P. Characterization of a buried neutral histidine residue in Bacillus circulans xylanase: NMR assignments, pH titration, and hydrogen exchange. *Protein Sci.* **5,** 2319–2328 (1996).

87. Goodall, J. J., Chen, G. J. & Page, M. G. P. Essential role of histidine 20 in the catalytic mechanism of Escherichia coli peptidyl-tRNA hydrolase. *Biochemistry* **43,** 4583–91 (2004).

88. Wiebe, C. a, Dibattista, E. R. & Fliegel, L. Functional role of polar amino acid residues in Na+/H+ exchangers. *Biochem. J.* **357,** 1–10 (2001).

89. Kalani, M. R., Moradi, A., Moradi, M. & Tajkhorshid, E. Characterizing a histidine switch controlling ph-dependent conformational changes of the influenza

virus hemagglutinin. *Biophys. J.* **105,** 993–1003 (2013).

90. McDonald, I. K. & Thornton, J. M. Satisfying Hydrogen Bonding Potential in Proteins. *J. Mol. Biol.* **238,** 777–793 (1994).

91. Tan, Y. J., Oliveberg, M., Davis, B. & Fersht, A. R. Perturbed pKA-values in the denatured states of proteins. *J. Mol. Biol.* **254,** 980–992 (1995).

92. Pace, C. N., Grimsley, G. R. & Scholtz, J. M. Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J. Biol. Chem.* **284,** 13285–9 (2009).

93. Krieger, E., Nielsen, J. E., Spronk, C. a E. M. & Vriend, G. Fast empirical pKa prediction by Ewald summation. *J. Mol. Graph. Model.* **25,** 481–486 (2006).

94. Kim, M. O., Nichols, S. E., Wang, Y. & McCammon, J. A. Effects of histidine protonation and rotameric states on virtual screening of M. tuberculosis RmlC. *J. Comput. Aided. Mol. Des.* **27,** 235–46 (2013).

95. Du, Z. *et al.* Highly Conserved Histidine Plays a Dual Catalytic Role in Protein Splicing: A p K a Shift Mechanism. *J. Am. Chem. Soc.* **131,** 11581–11589 (2009).

96. Li, C. *et al.* Metal-binding loop length is a determinant of the pKa of a histidine ligand at a type 1 copper site. *Inorg. Chem.* **50,** 482–8 (2011).

97. Edgcomb, S. P. & Murphy, K. P. Variability in the pKa of histidine side-chains correlates with burial within proteins. *Proteins* **49,** 1–6 (2002).

98. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).

99. Goddard, T. D. & Kneller, D. G. SPARKY 3. (2004).

100. Dwyer, J. J. *et al.* High apparent dielectric constants in the interior of a protein

reflect water penetration. *Biophys. J.* **79,** 1610–20 (2000).

101. Hirano, S., Kamikubo, H., Yamazaki, Y. & Kataoka, M. Elucidation of Information Encoded in Tryptophan 140 of. *Proteins Struct. Funct. Bioinforma.* **58,** 271–277 (2005).

102. Pace, C. N. & Shaw, K. L. Linear extrapolation method of analyzing solvent denaturation curves. *Proteins* **Suppl 4,** 1–7 (2000).

103. Huyghues-Despointes, B. M. P. *et al.* pK values of histidine residues in ribonuclease Sa: Effect of salt and net charge. *J. Mol. Biol.* **325,** 1093–1105 (2003).

104. Lee, K. K., Fitch, C. A. & García-Moreno E., B. Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein. *Protein Sci.* **11,** 1004–1016 (2002).

105. Kao, Y. H. *et al.* Salt effects on ionization equilibria of histidines in myoglobin. *Biophys. J.* **79,** 1637–1654 (2000).

106. Pelton, J. G., Torchia, D. A., Meadow, N. D. & Roseman, S. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Sci.* **2,** 543–558 (1993).

107. Vriend, G. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8,** 52–56, 29 (1990).

108. Karp, D. A., Stahley, M. R. & García-Moreno E., B. Conformational consequences of ionization of Lys, Asp, and Glu buried at position 66 in staphylococcal nuclease. *Biochemistry* **49,** 4138–46 (2010).

109. Chou, P. Y. & Fasman, G. D. Conformational parameters for amino acids in

helical, beta-sheet, and random coil regions calculated from proteins. *Biochemistry* **13,** 211–222 (1974).

110. Levitt, M. Conformational preferences of amino acids in globular proteins. *Biochemistry* **17,** 4277–4285 (1978).

111. Forsyth, W. R., Antosiewicz, J. M. & Robertson, A. D. Empirical relationships between protein structure and carboxyl pKa values in proteins. *Proteins Struct. Funct. Genet.* **48,** 388–403 (2002).

112. Harris, T. K. & Turner, G. J. Structural Basis of Perturbed pK a Values of Catalytic Groups in Enzyme Active Sites. *Life* **53,** 85–98 (2002).

113. Damjanovic, A., Brooks, B. R. & García-Moreno E., B. Conformational Relaxation and Water Penetration Coupled to Ionization of Internal Groups in Proteins. *Simulation* 4042–4053 (2011).

114. Goh, G. B., Laricheva, E. N. & Brooks, C. L. Uncovering pH-Dependent Transient Statese of Proteins with Buried Ionizable Residues. *J. Am. Chem. Soc.* **136,** 8496–8499 (2014).

115. Whitten, S. T., García-Moreno E., B. & Hilser, V. J. Local conformational fluctuations can modulate the coupling between proton binding and global structural transitions in proteins. *Proc. Natl. Acad. Sci.* **102,** 4282–4287 (2005).

116. Di Russo, N. V, Estrin, D. A., Martí, M. A. & Roitberg, A. E. pH-Dependent conformational changes in proteins and their effect on experimental pK(a)s: the case of Nitrophorin 4. *PLoS Comput. Biol.* **8,** e1002761 (2012).

117. Kumar, S., Ma, B., Tsai, C. J., Sinha, N. & Nussinov, R. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.* **9,** 10–19 (2000).

118. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5,** 789–796 (2009).

119. Gupta, S. & Bhattacharjya, S. NMR characterization of the near native and unfolded states of the PTB domain of DOK1: Alternate conformations and residual clusters. *PLoS One* **9,** (2014).

120. Bouvignies, G. *et al.* Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477,** 111–4 (2011).

121. Alexov, E. G. *et al.* Progress in the prediction of pKa values in proteins. *Proteins* **79,** 3260–75 (2011).

122. Robinson, A. C., Castañeda, C. A., Schlessman, J. L. & García-Moreno E., B. Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 11685–90 (2014).

123. Karp, D. A. *et al.* High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp. *Biophys. J.* **92,** 2041–2053 (2007).

124. Fitch, C. A. *et al.* Experimental pKa Values of Buried Residues: Analysis with Continuum Methods and Role of Water Penetration. *Biophys. J.* **82,** 3289–3304 (2002).

125. Harms, M. J. *et al.* A buried lysine that titrates with a normal pK_a: Role of conformational flexibility at the protein- water interface as a determinant of pK_a values. *Protein Sci.* **17,** 833–845 (2008).

126. Chen, J., Lu, Z., Sakon, J. & Stites, W. E. Increasing the thermostability of

staphylococcal nuclease: implications for the origin of protein thermostability. *J. Mol. Biol.* **303,** 125–30 (2000).

127. Hale, S. P., Poole, L. B. & Gerlt, J. A. Mechanism of the reaction catalyzed by staphylococcal nuclease: identification of the rate-determining step. *Biochemistry* **32,** 7479–87 (1993).

128. Whitten, S. T., Wooll, J. O., Razeghifard, R., García-Moreno E., B. & Hilser, V. J. The origin of pH-dependent changes in m-values for the denaturant-induced unfolding of proteins. *J. Mol. Biol.* **309,** 1165–75 (2001).

129. Lanyi, J. K. & Schobert, B. global conformational coupling in a heptahelical membrane protein: transport mechanism from crystal structures of the nine states in the bacteriorhodopsin photocycle. *Biochemistry* **43,** 3–8 (2004).

130. Fledderman, E. L. *et al.* Myristate exposure in the human immunodeficiency virus type 1 matrix protein is modulated by pH. *Biochemistry* **49,** 9551–62 (2010).

131. Williams, S. L., De Oliveira, C. A. F. & J. Andrew McCammon, . Coupling constant pH molecular dynamics with accelerated molecular dynamics. *J. Chem. Theory Comput.* **6,** 560–568 (2010).

132. Itoh, S. G., Damjanović, A. & Brooks, B. R. PH replica-exchange method based on discrete protonation states. *Proteins Struct. Funct. Bioinforma.* **79,** 3420–3436 (2011).

133. Chen, W., Morrow, B. H., Shi, C. & Shen, J. K. Recent development and application of constant pH molecular dynamics. *Mol. Simul.* **40,** 830–838 (2014).

134. Witham, S. *et al.* Developing hybrid approaches to predict pK a values of ionizable groups. *Proteins Struct. Funct. Bioinforma.* **79,** 3389–3399 (2011).

135. JKM, R., MA, H., AP, M., Edwards, S. & Webster, C. Contribution of malate and amino acid metabolism to cytoplasmic pH regulation in hypoxic maize root tips studied using nuclear magnetic resonance spectroscopy. *Plant Physiol.* **98,** 480 (1992).

136. Sakano, K. Metabolic regulation of pH in plant cells: Role of cytoplasmic pH in defense reaction and secondary metabolism. *Int. Rev. Cytol.* **206,** 1–44 (2001).

137. Yao, H. & Haddad, G. G. Calcium and pH homeostasis in neurons during hypoxia and ischemia. *Cell Calcium* **36,** 247–255 (2004).

138. Jentsch, T. J. *CLC Chloride Channels and Transporters: From Genes to Protein Structure, Pathology and Physiology*. *Crit. Rev. Biochem. Mol. Biol.* **43,** (2008).

139. Scott, C. C. & Gruenberg, J. Ion flux and the function of endosomes and lysosomes: PH is just the start: The flux of ions across endosomal membranes influences endosome function not only through regulation of the luminal pH. *BioEssays* **33,** 103–110 (2011).

140. Stevens, T. H. & Forgac, M. Structure, function and regulation of the vacuolar (H+)-ATPase. *Annu. Rev. Cell Dev. Biol.* **13,** 779–808 (1997).

141. Falhof, J., Pedersen, J. T., Fuglsang, A. T. & Palmgren, M. Plasma membrane H+-ATPase regulation in the center of plant physiology. *Mol. Plant* **9,** 323–337 (2015).

142. Boron, W. F. Intracellular pH regulation in epithelial cells. *Annu. Rev. Physiol.* **48,** 377–388 (1986).

143. Frantz, C. *et al.* Cofilin is a pH sensor for actin free barbed end formation: role of phosphoinositide binding. *J. Cell Biol.* **183,** 865–79 (2008).

144. Lagadic-Gossmann, D., Huc, L. & Lecureur, V. Alterations of intracellular pH

homeostasis in apoptosis: origins and roles. *Cell Death Differ.* **11,** 953–61 (2004).

145. Kornak, U. *et al.* Impaired glycosylation and cutis laxa caused by mutations in the vesicular H+-ATPase subunit ATP6V0A2. *Nat. Genet.* **40,** 32–4 (2008).

146. Casey, J. R., Grinstein, S. & Orlowski, J. Sensors and regulators of intracellular pH. *Nat. Rev. Mol. Cell Biol.* **11,** 50–61 (2010).

147. Obara, M., Szeliga, M. & Albrecht, J. Regulation of pH in the mammalian central nervous system under normal and pathological conditions: Facts and hypotheses. *Neurochem. Int.* **52,** 905–919 (2008).

148. Vaughan-Jones, R. D., Spitzer, K. W. & Swietach, P. Intracellular pH regulation in heart. *J. Mol. Cell. Cardiol.* **46,** 318–331 (2009).

149. Coffey, E. E., Beckel, J. M., Laties, A. M. & Mitchell, C. H. Lysosomal alkalization and dysfunction in human fibroblasts with the alzheimer's disease-linked presenilin 1 A246E mutation can be reversed with cAMP. *Neuroscience* **263,** 111–124 (2014).

150. Buell, A. K. *et al.* Solution conditions determine the relative importance of nucleation and growth processes in α-synuclein aggregation. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 7671–7676 (2014).

151. Swietnicki, W., Petersen, R., Gambetti, P. & Surewicz, W. K. pH-dependent Stability and Conformation of the Recombinant Human Prion Protein PRP(90-231). *J. Biol. Chem.* **272,** 27517–27520 (1997).

152. Ihara, Y. *et al.* The G protein-coupled receptor T-cell death-associated gene 8 (TDAG8) facilitates tumor development by serving as an extracellular pH sensor. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 17309–14 (2010).

153. Webb, B. a, Chimenti, M. S., Jacobson, M. P. & Barber, D. L. Dysregulated pH: a perfect storm for cancer progression. *Nat. Rev. Cancer* **11,** 671–7 (2011).

154. Rivinoja, A., Pujol, F. M., Hassinen, A. & Kellokumpu, S. Golgi pH, its regulation and roles in human disease. *Ann. Med.* **44,** 542–54 (2012).

155. Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371,** 37–43 (1994).

156. Li, W., Nicol, F. & Szoka, F. C. GALA: a designed synthetic pH-responsive amphipathic peptide with applications in drug and gene delivery. *Adv. Drug Deliv. Rev.* **56,** 967–85 (2004).

157. Reshetnyak, Y. K., Andreev, O. A., Lehnert, U. & Engelman, D. M. Translocation of molecules into cells by pH-dependent insertion of a transmembrane helix. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 6460–5 (2006).

158. Stratton, M. M. & Loh, S. N. On the mechanism of protein fold-switching by a molecular sensor. *Proteins* **78,** 3260–9 (2010).

159. Ostermeier, M. Engineering allosteric protein switches by domain insertion. *Protein Eng. Des. Sel.* **18,** 359–64 (2005).

160. Watanabe, H. *et al.* Optimizing pH response of affinity between protein G and IgG Fc. How electrostatic modulations affect protein-protein interactions. *J. Biol. Chem.* **284,** 12373–12383 (2009).

161. Strauch, E., Fleishman, S. J. & Baker, D. Computational design of a pH-sensitive IgG binding protein. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 675–680 (2014).

162. Bell-Upp, P. *et al.* Thermodynamic principles for the engineering of pH-driven

conformational switches and acid insensitive proteins. *Biophys. Chem.* **159,** 217–26 (2011).

163. Manson, A., Whitten, S. T., Ferreon, J. C., Fox, R. O. & Hilser, V. J. Characterizing the role of ensemble modulation in mutation-induced changes in binding affinity. *J. Am. Chem. Soc.* **131,** 6785–93 (2009).

164. Boron, W. F. Regulation of intracellular pH. *Adv. Physiol. Educ.* **28,** 160–79 (2004).

165. Robergs, R. A., Ghiasvand, F. & Parker, D. Biochemistry of exercise-induced metabolic acidosis. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **287,** R502–R516 (2004).

166. Xiong, Z. G. *et al.* Neuroprotection in ischemia: Blocking calcium-permeable acid-sensing ion channels. *Cell* **118,** 687–698 (2004).

167. Warburg, O. On the Origin of Cancer Cells. *Science (80-.).* **123,** 309–314 (1956).

168. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324,** 1029–33 (2009).

169. Riggs, A. F. The Bohr effect. *Annu. Rev. Physiol.* **50,** 181–204 (1988).

170. Skehel, J. J. & Wiley, D. C. RECEPTOR BINDING AND MEMBRANE FUSION IN VIRUS ENTRY: The Influenza Hemagglutinin. *Annu. Rev. Biochem.* **69,** 531–569 (2000).

171. Fritz, R., Stiasny, K. & Heinz, F. X. Identification of specific histidines as pH sensors in flavivirus membrane fusion. *J. Cell Biol.* **183,** 353–61 (2008).

172. Harrison, J. S. *et al.* Role of electrostatic repulsion in controlling pH-dependent

conformational changes of viral fusion proteins. *Structure* **21,** 1085–96 (2013).

173. Burdick, D. *et al.* Assembly and aggregation properties of synthetic Alzheimer's A4/beta amyloid peptide analogs. *J. Biol. Chem.* **267,** 546–554 (1992).

174. Jha, S. *et al.* PH dependence of amylin fibrillization. *Biochemistry* **53,** 300–310 (2014).

175. Srivastava, J., Barber, D. L. & Jacobson, M. P. Intracellular pH sensors: design principles and functional significance. *Physiology (Bethesda).* **22,** 30–9 (2007).

176. Ambroggio, X. I. & Kuhlman, B. Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J. Am. Chem. Soc.* **128,** 1154–61 (2006).

177. Lee, H. M. & Chmielewski, J. Liposomal cargo unloading induced by pH-sensitive peptides. *J. Pept. Res.* **65,** 355–363 (2005).

178. Schröter, C. *et al.* A generic approach to engineer antibody pH-switches using combinatorial histidine scanning libraries and yeast display. *MAbs* **7,** 138–151 (2015).

179. Hanakam, F., Albrecht, R., Eckerskorn, C., Matzner, M. & Gerisch, G. Myristoylated and non-myristoylated forms of the pH sensor protein hisactophilin II: intracellular shuttling to plasma membrane and nucleus monitored in real time by a fusion with green fluorescent protein. *EMBO J.* **15,** 2935–43 (1996).

180. Wachter, R. M., Elsliger, M. A., Kallio, K., Hanson, G. T. & Remington, S. J. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure* **6,** 1267–77 (1998).

181. Bagar, T., Altenbach, K., Read, N. D. & Bencina, M. Live-Cell imaging and

measurement of intracellular pH in filamentous fungi using a genetically encoded ratiometric probe. *Eukaryot. Cell* **8,** 703–12 (2009).

182. Orij, R., Postmus, J., Ter Beek, A., Brul, S. & Smits, G. J. In vivo measurement of cytosolic and mitochondrial pH using a pH-sensitive GFP derivative in Saccharomyces cerevisiae reveals a relation between intracellular pH and growth. *Microbiology* **155,** 268–78 (2009).

183. Shen, J. *et al.* Organelle pH in the Arabidopsis endomembrane system. *Mol. Plant* **6,** 1419–37 (2013).

184. Martinière, A. *et al.* In vivo intracellular pH measurements in tobacco and Arabidopsis reveal an unexpected pH gradient in the endomembrane system. *Plant Cell* **25,** 4028–43 (2013).

185. Bizzarri, R. *et al.* Development of a novel GFP-based ratiometric excitation and emission pH indicator for intracellular studies. *Biophys. J.* **90,** 3300–14 (2006).

186. Valkonen, M., Mojzita, D., Penttilä, M. & Bencina, M. Noninvasive high-throughput single-cell analysis of the intracellular pH of Saccharomyces cerevisiae by ratiometric flow cytometry. *Appl. Environ. Microbiol.* **79,** 7179–87 (2013).

187. Maresova, L., Haskova, B., Urbankova, E., Chaloupka, R. & Sychrova, H. New applications of pHluorin — measuring intracellular pH of prototrophic yeasts and determining changes in the buffering capacity of strains with affected potassium homeostasis. *Yeast* **27,** 317–325 (2010).

188. Poëa-Guyon, S., Pasquier, H., Mérola, F., Morel, N. & Erard, M. The enhanced cyan fluorescent protein: a sensitive pH sensor for fluorescence lifetime imaging. *Anal. Bioanal. Chem.* **405,** 3983–7 (2013).

189. Stratton, M. M. & Loh, S. N. Converting a protein into a switch for biosensing and functional regulation. *Protein Sci.* **20,** 19–29 (2011).

190. Wright, C. M., Wright, R. C., Eshleman, J. R. & Ostermeier, M. A protein therapeutic modality founded on molecular regulation. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 16206–11 (2011).

191. Sagermann, M., Chapleau, R. R., DeLorimier, E. & Lei, M. Using affinity chromatography to engineer and characterize pH-dependent protein switches. *Protein Sci.* **18,** 217–28 (2009).

192. Berbasova, T. *et al.* Rational design of a colorimetric pH sensor from a soluble retinoic acid chaperone. *J. Am. Chem. Soc.* **135,** 16111–9 (2013).

193. Igawa, T. *et al.* Antibody recycling by engineered pH-dependent antigen binding improves the duration of antigen neutralization. *Nat. Biotechnol.* **28,** 1203–7 (2010).

194. Chaparro-Riggers, J. *et al.* Increasing serum half-life and extending cholesterol lowering in vivo by engineering antibody with pH-sensitive binding to PCSK9. *J. Biol. Chem.* **287,** 11090–7 (2012).

195. Shortle, D., Meeker, a K. & Freire, E. Stability mutants of staphylococcal nuclease: large compensating enthalpy-entropy changes for the reversible denaturation reaction. *Biochemistry* **27,** 4761–4768 (1988).

Appendix

Table A1 Data collection and refinement statistics for internal polar variants of staphylococcal nuclease.

	Δ+PHS V23S	**Δ+PHS I92S**
Crystallization conditions		
Temperature, K	277	277
pH	9	8
Buffer	25 mM KH_2PO_4/K_2HPO_4	25 mM KH_2PO_4/K_2HPO_4
Precipitant	26 % (w/v) MPD	25 % (w/v) MPD
Additives‡	2 *eq.* pdTp, 3 *eq.* $CaCl_2$	2 *eq.* pdTp, 3 *eq.* $CaCl_2$
Data collection statistics		
Space group	P 1 2_1 1	P 1 2_1 1
Unit cell dimensions		
a, Å	31.15	31.03
b, Å	60.40	59.96
c, Å	37.91	38.02
α, °	90.00	90.00
β, °	93.68	93.63
γ, °	90.00	90.00
Temperature , K	100	110
Radiation source	NSLS X25	Bruker ApexII DUO
Wavelength, Å	1.10	1.54
Resolution range, Å	50.0 – 1.60 (1.63 – 1.60)*	50.00 – 1.80 (1.83 - 1.80)
No. of unique reflections	18541	12978
Completeness, %	99.7 (97.4)	99.9 (100.0)
R_{merge}¶ or R_{sigma}¥	0.054 (0.217)	0.035 (0.227)
Redundancy	6.4	9.1
Average $I/\sigma(I)$	46.05	18.95
Wilson B, Å^2	30.1	27.2
Refinement statistics		
Resolution range, Å	32.06 – 1.60 (1.64 – 1.60)	37.94 – 1.80 (1.85 – 1.80)
No. of reflections	18521 (1367)	12950 (952)
Completeness, %	99.6 (96.9)	99.8 (99.8)
R_{factor}	0.171 (0.233)	0.198 (0.262)
R_{free}	0.207 (0.299)	0.227 (0.295)
No. of non-hydrogen atoms	1207	1223
Protein	1063	1130
Water	118	67
Ion / Ligand	26	26
RMSD from ideal geometry		
Bonds, Å	0.016	0.018
Angles, °	1.74	1.82
Average B factors, Å^2		
Protein	27.7	21.8
Water	32.2	23.0
Ion / Ligand	19.8	16.8
Ramachandran plot		

No. in favored region (%)	114 (94.2)	112 (92.6)
No. in allowed region (%)	6 (5.0)	8 (6.6)
No. in outlier region (%)	1 (0.8)	1 (0.8)
PDB ID code	**4KHV**	**4PMB**

	Δ+PHS L25Y	Δ+PHS I92T
Crystallization conditions		
Temperature, K	277	277
pH	7	9
Buffer	25 mM KH_2PO_4/K_2HPO_4	25 mM KH_2PO_4/K_2HPO_4
Precipitant	30 % (w/v) MPD	35 % (w/v) MPD
Additives‡	1 *eq.* pdTp, 2 *eq.* $CaCl_2$	2 *eq.* pdTp, 3 *eq.* $CaCl_2$
Data collection statistics		
Space group	P 1 2_1 1	P 1 2_1 1
Unit cell dimensions		
a, Å	30.89	31.08
b, Å	60.13	60.22
c, Å	38.14	38.04
α, °	90.00	90.00
β, °	93.72	93.03
γ, °	90.00	90.00
Temperature , K	110	110
Radiation source	Bruker ApexII DUO	Bruker ApexII DUO
Wavelength, Å	1.54	1.54
Resolution range, Å	50.0 – 2.10 (2.15 – 2.10)*	50.00 – 1.90 (1.93 - 1.90)
No. of unique reflections	8185	11102
Completeness, %	99.5 (99.1)	99.7 (98.2)
R_{merge}¶ or R_{sigma}¥	0.044 (0.243)	0.030 (0.186)
Redundancy	9.2	10.1
Average $I/\sigma(I)$	15.99	21.60
Wilson B, Å^2	37.3	30.9
Refinement statistics		
Resolution range, Å	32.06 – 1.60 (1.64 – 1.60)	32.13 – 1.90 (1.95 – 1.90)
No. of reflections	18521 (1367)	10565 (804)
Completeness, %	99.6 (96.9)	99.7 (98.2)
R_{factor}	0.171 (0.233)	0.190 (0.265)
R_{free}	0.207 (0.299)	0.253 (0.355)
No. of non-hydrogen atoms	1207	1142
Protein	1063	1032
Water	118	84
Ion / Ligand	26	26
RMSD from ideal geometry		
Bonds, Å	0.016	0.018
Angles, °	1.74	1.81
Average B factors, Å^2		
Protein	27.7	27.8
Water	32.2	29.6
Ion / Ligand	19.8	22.3
Ramachandran plot		
No. in favored region (%)	114 (94.2)	112 (92.6)
No. in allowed region (%)	6 (5.0)	8 (6.6)
No. in outlier region (%)	1 (0.8)	1 (0.8)
PDB ID code	**XXXX**	**4ZQ3**

‡eq. = molar equivalents (relative to 1 molar equivalent of protein)
*Values in parentheses correspond to the highest resolution shell.

¶ $R_{merge} = \frac{\sum_{hkl}\sum_{j}\left|I_{hkl,j} - \langle I_{hkl}\rangle\right|}{\sum_{hkl}\sum_{j} I_{hkl,j}}$; ¥ $R_{sigma} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}\left|I_{hkl,j} - \langle I_{hkl}\rangle\right|}{\sum_{hkl}\sum_{j} I_{hkl,j}}$

Table A2 Data collection and refinement statistics for internal histidine variants of staphylococcal nuclease.

	Δ+PHS V23H	**Δ+PHS L25H**
Crystallization conditions		
Temperature, K	277	277
pH	7	9
Buffer	25 mM KH_2PO_4/K_2HPO_4	25 mM KH_2PO_4/K_2HPO_4
Precipitant	30 % (w/v) MPD	30 % (w/v) MPD
Additives‡	2 *eq.* pdTp, 3 *eq.* $CaCl_2$	2 *eq.* pdTp, 3 *eq.* $CaCl_2$
Data collection statistics		
Space group	P 1 2_1 1	P 1 2_1 1
Unit cell dimensions		
a, Å	31.10	31.04
b, Å	60.42	60.51
c, Å	38.08	38.11
α, °	90.00	90.00
β, °	94.41	93.95
γ, °	90.00	90.00
Temperature , K	110	110
Radiation source	Bruker ApexII DUO	Bruker ApexII DUO
Wavelength, Å	1.54	1.54
Resolution range, Å	50.00 – 1.70 (1.72 – 1.70)*	50.00 – 1.62 (1.64 - 1.62)
No. of unique reflections	15518	17960
Completeness, %	99.9 (100.0)	100.0 (100.0)
R_{merge}¶ or R_{sigma}¥	0.025 (0.172)	0.021 (0.176)
Redundancy	13.21	11.79
Average $I/\sigma(I)$	29.4	32.1
Wilson B, $Å^2$	22.8	22.8
Refinement statistics		
Resolution range, Å	32.15 – 1.70 (1.74 – 1.70)	32.19 – 1.62 (1.66 – 1.62)
No. of reflections	14717 (1144)	17023 (1244)
Completeness, %	99.8 (99.9)	100.0 (100.0)
R_{factor}	0.213 (0.239)	0.169 (0.185)
R_{free}	0.251 (0.292)	0.192 (0.235)
No. of non-hydrogen atoms	1157	1263
Protein	1036	1111
Water	95	126
Ion / Ligand	26	26
RMSD from ideal geometry		
Bonds, Å	0.017	0.019
Angles, °	1.83	1.93
Average B factors, $Å^2$		
Protein	18.0	17.1
Water	19.8	23.7
Ion / Ligand	18.7	22.4
Ramachandran plot		
No. in favored region (%)	103 (93.6)	106 (93.8)
No. in allowed region (%)	6 (5.5)	6 (5.3)
No. in outlier region (%)	1 (0.9)	1 (0.9)
PDB ID code	**4ZUI**	**5C4Z**

	Δ+PHS L36H	Δ+PHS T62H
Crystallization conditions		
Temperature, K	277	277
pH	8	8
Buffer	25 mM KH_2PO_4/K_2HPO_4	25 mM KH_2PO_4/K_2HPO_4
Precipitant	25 % (w/v) MPD	19 % (w/v) MPD
Additives‡	2 *eq.* pdTp, 3 *eq.* $CaCl_2$	2 *eq.* pdTp, 3 *eq.* $CaCl_2$
Data collection statistics		
Space group	P 1 2_1 1	P 1 2_1 1
Unit cell dimensions		
a, Å	31.05	31.17
b, Å	60.33	60.72
c, Å	38.20	38.16
α, °	90.00	90.00
β, °	93.31	93.47
γ, °	90.00	90.00
Temperature , K	100	110
Radiation source	NSLS X25	SuperNova (Cu)
Wavelength, Å	1.10	1.54
Resolution range, Å	50.0 – 1.58 (1.61 – 1.58)*	50.00 – 1.85 (1.92 - 1.85)
No. of unique reflections	19399	12220
Completeness, %	100.0 (99.9)	99.9 (100.0)
R_{merge}¶ or R_{sigma}¥	0.059 (0.308)	0.020 (0.118)
Redundancy	6.6	9.9
Average $I/\sigma(I)$	11.2	6.52
Wilson B, Å^2	28.6	25.3
Refinement statistics		
Resolution range, Å	38.14– 1.58 (1.62 – 1.58)	38.09 – 1.85 (1.90 – 1.85)
No. of reflections	19334 (1379)	11531 (901)
Completeness, %	100.0 (99.9)	99.9 (100.0)
R_{factor}	0.173 (0.202)	0.189 (0.258)
R_{free}	0.191 (0.267)	0.228 (0.305)
No. of non-hydrogen atoms	1150	1157
Protein	1035	1036
Water	89	95
Ion / Ligand	26	26
RMSD from ideal geometry		
Bonds, Å	0.016	0.018
Angles, °	1.98	1.81
Average B factors, Å^2		
Protein	23.5	18.6
Water	29.3	20.1
Ion / Ligand	27.9	12.4
Ramachandran plot		
No. in favored region (%)	111 (94.1)	116 (93.6)
No. in allowed region (%)	6 (5.0)	7 (5.7)
No. in outlier region (%)	1 (0.9)	1 (0.8)
PDB ID code	**4LAA**	**5I9P**

	Δ+PHS V66H	**Δ+PHS I72H**
Crystallization conditions		
Temperature, K	277	277
pH	6	8
Buffer	25 mM KH_2PO_4/K_2HPO_4	25 mM KH_2PO_4/K_2HPO_4
Precipitant	20 % (w/v) MPD	20 % (w/v) MPD
Additives‡	2 *eq.* pdTp, 3 *eq.* $CaCl_2$	1 *eq.* pdTp, 2 *eq.* $CaCl_2$
Data collection statistics		
Space group	P 1 2_1 1	P 1 2_1 1
Unit cell dimensions		
a, Å	60.68	31.17
b, Å	76.86	60.57
c, Å	31.25	37.44
α, °	90.00	90.00
β, °	90.00	93.79
γ, °	90.00	90.00
Temperature , K	110	110
Radiation source	Bruker ApexII DUO	Bruker ApexII DUO
Wavelength, Å	1.54	1.54
Resolution range, Å	50.0 – 1.72 (1.74 – 1.72)*	50.00 – 1.68 (1.71 - 1.68)
No. of unique reflections	14985	15921
Completeness, %	100.0 (100.0)	99.9 (100.0)
R_{merge}¶ or R_{sigma}¥	0.025 (0.176)	0.030 (0.209)
Redundancy	12.7	8.0
Average $I/\sigma(I)$	27.39	20.59
Wilson B, Å^2	22.6	25.2
Refinement statistics		
Resolution range, Å	32.17 – 1.72 (1.77 – 1.72)	31.8 – 1.68 (1.72 – 1.68)
No. of reflections	14207 (1037)	15882 (1180)
Completeness, %	100.0 (100.0)	99.8 (100.0)
R_{factor}	0.166 (0.189)	0.198 (0.250)
R_{free}	0.207 (0.225)	0.242 (0.279)
No. of non-hydrogen atoms	1231	1223
Protein	1097	1044
Water	108	86
Ion / Ligand	26	26
RMSD from ideal geometry		
Bonds, Å	0.018	0.016
Angles, °	1.81	1.90
Average B factors, Å^2		
Protein	19.7	23.2
Water	25.0	24.8
Ion / Ligand	23.8	23.8
Ramachandran plot		
No. in favored region (%)	108 (94.7)	110 (92.4)
No. in allowed region (%)	5 (4.4)	8 (6.7)
No. in outlier region (%)	1 (0.9)	1 (0.8)
PDB ID code	**5C3W**	**4PNY**

	Δ+PHS Y91H	**Δ+PHS I92H**
Crystallization conditions		
Temperature, K	277	277
pH	6	6
Buffer	25 mM KH_2PO_4/K_2HPO_4	25 mM KH_2PO_4/K_2HPO_4
Precipitant	21 % (w/v) MPD	30 % (w/v) MPD
Additives‡	1 *eq.* pdTp, 2 *eq.* $CaCl_2$	2 *eq.* pdTp, 3 *eq.* $CaCl_2$
Data collection statistics		
Space group	P 1 2_1 1	P 1 2_1 1
Unit cell dimensions		
a, Å	31.09	31.01
b, Å	60.46	60.28
c, Å	38.26	38.10
α, °	90.00	90.00
β, °	93.64	94.09
γ, °	90.00	90.00
Temperature , K	110	110
Radiation source	Bruker ApexII DUO	Bruker ApexII DUO
Wavelength, Å	1.54	1.54
Resolution range, Å	50.0 – 1.72 (1.74 – 1.72)*	60.28 – 1.80 (1.83 - 1.80)
No. of unique reflections	15097	13063
Completeness, %	99.9 (99.8)	100.0 (100.0)
R_{merge}¶ or R_{sigma}¥	0.028 (0.176)	0.018 (0.161)
Redundancy	3.8	13.8
Average $I/\sigma(I)$	22.26	35.50
Wilson B, Å^2	26.4	29.6
Refinement statistics		
Resolution range, Å	31.03 – 1.72 (1.77 – 1.72)	32.15 – 1.80 (1.85 – 1.80)
No. of reflections	14316 (1098)	12375 (915)
Completeness, %	99.9 (99.9)	99.7 (100.0)
R_{factor}	0.175 (0.187)	0.176 (0.244)
R_{free}	0.207 (0.226)	0.196 (0.292)
No. of non-hydrogen atoms	1160	1202
Protein	1031	1084
Water	103	92
Ion / Ligand	26	26
RMSD from ideal geometry		
Bonds, Å	0.018	0.019
Angles, °	2.00	1.79
Average B factors, Å^2		
Protein	22.1	25.9
Water	26.7	29.9
Ion / Ligand	25.5	28.5
Ramachandran plot		
No. in favored region (%)	106 (93.8)	111 (94.1)
No. in allowed region (%)	6 (5.3)	6 (5.1)
No. in outlier region (%)	1 (0.9)	1 (0.9)
PDB ID code	**4ZUJ**	**5C4H**

Table A3. Data collection and refinement statistics for Δ+VIAGLA V23K.

	Δ+VIAGLA V23K
Crystallization conditions	
Temperature, K	277
pH	8
Buffer	25 mM KH_2PO_4/K_2HPO_4
Precipitant	25 % (w/v) MPD
Additives‡	2 *eq.* pdTp, 3 *eq.* $CaCl_2$
Data collection statistics	
Space group	P 2_1 2_1 2_1
Unit cell dimensions	
a, Å	30.93
b, Å	60.77
c, Å	78.99
α, °	90.00
β, °	90.00
γ, °	90.00
Temperature , K	110
Radiation source	Bruker ApexII DUO
Wavelength, Å	1.54
Resolution range, Å	50.00 – 1.75 (1.78 – 1.75)*
No. of unique reflections	15697
Completeness, %	100.0 (100.0)
R_{merge}¶ or R_{sigma}¥	0.019 (0.120)
Redundancy	15.8
Average $I/\sigma(I)$	35.1
Wilson B, Å^2	26.5
Refinement statistics	
Resolution range, Å	48.17 – 1.75 (1.80 – 1.75)
No. of reflections	15640 (1139)
Completeness, %	99.9 (100.0)
R_{factor}	0.208 (0.226)
R_{free}	0.261 (0.271)
No. of non-hydrogen atoms	1202
Protein	1066
Water	103
Ion / Ligand	33
RMSD from ideal geometry	
Bonds, Å	0.017
Angles, °	1.83
Average B factors, Å^2	
Protein	14.0
Water	24.5
Ion / Ligand	20.4
Ramachandran plot	
No. in favored region (%)	115 (95.0)
No. in allowed region (%)	5 (4.1)
No. in outlier region (%)	1 (0.8)
PDB ID code	**XXXX**

‡eq. = molar equivalents (relative to 1 molar equivalent of protein)
*Values in parentheses correspond to the highest resolution shell.

¶ $R_{merge} = \frac{\sum_{hkl}\sum_{j}\left|I_{hkl,j} - \langle I_{hkl}\rangle\right|}{\sum_{hkl}\sum_{j} I_{hkl,j}}$; ¥ $R_{sigma} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}\left|I_{hkl,j} - \langle I_{hkl}\rangle\right|}{\sum_{hkl}\sum_{j} I_{hkl,j}}$

Table A4. Data collection and refinement statistics for Δ+PHS V66E A109E.

	Δ+PHS V66E A09E
Crystallization conditions	
Temperature, K	277
pH	6
Buffer	25 mM KH_2PO_4/K_2HPO_4
Precipitant	30 % (w/v) MPD
Additives‡	1 *eq.* pdTp, 2 *eq.* $CaCl_2$
Data collection statistics	
Space group	P 2_1 2_1 2_1
Unit cell dimensions	
a, Å	47.15
b, Å	54.16
c, Å	108.13
α, °	90.00
β, °	90.00
γ, °	90.00
Temperature , K	110
Radiation source	Bruker ApexII DUO
Wavelength, Å	1.54
Resolution range, Å	50.00 – 1.67 (1.69 – 1.67)*
No. of unique reflections	32967
Completeness, %	99.8 (100.0)
R_{merge}¶ or R_{sigma}¥	0.026 (0.228)
Redundancy	13.0
Average $I/\sigma(I)$	29.3
Wilson B, Å^2	22.1
Refinement statistics	
Resolution range, Å	48.47 – 1.67 (1.71 – 1.67)
No. of reflections in test set	1669 (148)
Completeness, %	99.8 (100.0)
R_{work}	0.158 (0.181)
R_{free}	0.191 (0.249)
No. of non-hydrogen atoms	2639
Protein	2297
Water	290
Ion / Ligand	52
RMSD from ideal geometry	
Bonds, Å	0.016
Angles, °	1.80
Average B factors, Å^2	
Protein	14.0
Water	23.1
Ion / Ligand	10.7
Ramachandran plot	
No. in favored region (%)	113 (87.8)
No. in allowed region (%)	7 (11.3)
No. in outlier region (%)	1 (0.8)

PDB ID code **4OL7**

‡eq. = molar equivalents (relative to 1 molar equivalent of protein)
*Values in parentheses correspond to the highest resolution shell.

¶ $R_{merge} = \frac{\sum_{hkl}\sum_{j}\left|I_{hkl,j} - \langle I_{hkl}\rangle\right|}{\sum_{hkl}\sum_{j} I_{hkl,j}}$; ¥ $R_{sigma} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}\left|I_{hkl,j} - \langle I_{hkl}\rangle\right|}{\sum_{hkl}\sum_{j} I_{hkl,j}}$

Table A5. Data collection and refinement statistics for double Lys variants of Δ+PHS

	Δ+PHS L36K L103K	**Δ+PHS L36K L103K**
Crystallization conditions		
Temperature, K	277	277
pH	6	7
Buffer	25 mM KH_2PO_4/K_2HPO_4	25 mM KH_2PO_4/K_2HPO_4
Precipitant	25 % (w/v) MPD	20 % (w/v) MPD
Additives‡	2 *eq.* pdTp, 3 *eq.* $CaCl_2$	2 *eq.* pdTp, 3 *eq.* $CaCl_2$
Data collection statistics		
Space group	P 2_1 2_1 2_1	P 1 2_1 1
Unit cell dimensions		
a, Å	52.56	31.19
b, Å	53.27	78.93
c, Å	106.89	60.89
α, °	90.00	90.00
β, °	90.00	90.18
γ, °	90.00	90.00
Temperature , K	110	110
Radiation source	Bruker ApexII DUO	SuperNova (Cu)
Wavelength, Å	1.54	1.54
Resolution range, Å	50.0 – 2.05 (2.15 – 2.05)*	60.89 – 1.95 (2.02 - 1.95)
No. of unique reflections	19521	21550
Completeness, %	100.0 (100.0)	99.9 (99.9)
R_{merge}¶ or R_{sigma}¥	0.027 (0.238)	0.047 (0.216)
Redundancy	24.3	9.7
Average $I/\sigma(I)$	28.75	28.05
Wilson B, Å^2	36.5	22.8
Refinement statistics		
Resolution range, Å	37.73 – 2.05 (2.10 – 2.05)	60.89 – 1.95 (2.00 – 1.95)
No. of reflections	18445 (1414)	20492 (1486)
Completeness, %	99.8 (100.0)	99.9 (99.9)
R_{factor}	0.254 (0.336)	0.202 (0.276)
R_{free}	0.311 (0.418)	0.238 (0.326)
No. of non-hydrogen atoms	2201	2287
Protein	2103	2078
Water	46	158
Ion / Ligand	52	52
RMSD from ideal geometry		
Bonds, Å	0.015	0.017
Angles, °	1.88	1.81
Average B factors, Å^2		
Protein	32.1	15.8
Water	28.2	18.2
Ion / Ligand	27.7	10.2
Ramachandran plot		
No. in favored region (%)	248 (95.8)	239 (94.8)
No. in allowed region (%)	9 (3.5)	11 (4.4)
No. in outlier region (%)	2 (0.8)	2 (0.8)
PDB ID code	**xxxx**	**5I9O**

	Δ+PHS V66K A109K	Δ+PHS I72K V74K
Crystallization conditions		
Temperature, K	277	277
pH	9	7
Buffer	25 mM KH_2PO_4/K_2HPO_4	25 mM KH_2PO_4/K_2HPO_4
Precipitant	21 % (w/v) MPD	25 % (w/v) MPD
Additives‡	2 *eq.* pdTp, 3 *eq.* $CaCl_2$	2 *eq.* pdTp, 3 *eq.* $CaCl_2$
Data collection statistics		
Space group	P 1 2_1 1	P 2_1 2_1 2_1
Unit cell dimensions		
a, Å	31.03	31.88
b, Å	61.16	59.89
c, Å	38.91	73.76
α, °	90.00	90.00
β, °	92.24	90.00
γ, °	90.00	90.00
Temperature , K	100	110
Radiation source	NSLS X25	Bruker ApexII DUO
Wavelength, Å	1.10	1.54
Resolution range, Å	50.00 – 1.45 (1.48 - 1.45)	50.0 – 1.65 (1.75 – 1.65)*
No. of unique reflections	25640	17697
Completeness, %	99.5 (96.7)	99.9 (99.6)
R_{merge}¶ or R_{sigma}¥	0.043 (0.314)	0.019 (0.189)
Redundancy	6.3	35.4
Average $I/\sigma(I)$	52.4	47.22
Wilson B, Å^2	28.8	21.6
Refinement statistics		
Resolution range, Å	32.81 – 1.45 (1.49 – 1.45)	36.88 – 1.65 (1.69 – 1.65)
No. of reflections	24318 (1808)	16746 (1278)
Completeness, %	99.4 (95.3)	99.9 (99.2)
R_{factor}	0.203 (0.306)	0.183 (0.267)
R_{free}	0.233 (0.311)	0.210 (0.291)
No. of non-hydrogen atoms	1233	1182
Protein	1101	1056
Water	83	100
Ion / Ligand	26	26
RMSD from ideal geometry		
Bonds, Å	0.015	0.017
Angles, °	1.76	1.78
Average B factors, Å^2		
Protein	26.6	14.3
Water	32.4	20.5
Ion / Ligand	23.7	12.9
Ramachandran plot		
No. in favored region (%)	109 (94.0)	115 (93.5)
No. in allowed region (%)	6 (5.2)	7 (5.7)
No. in outlier region (%)	1 (0.9)	1 (0.8)
PDB ID code	**xxxx**	**4PMC**

	Δ+PHS I72K V74K
Crystallization conditions	
Temperature, K	277
pH	9
Buffer	25 mM KH_2PO_4/K_2HPO_4
Precipitant	20 % (w/v) MPD
Additives‡	2 *eq.* pdTp, 3 *eq.* $CaCl_2$
Data collection statistics	
Space group	P 2_1 2_1 2_1
Unit cell dimensions	
a, Å	31.81
b, Å	60.20
c, Å	73.83
α, °	90.00
β, °	90.00
γ, °	90.00
Temperature , K	110
Radiation source	Bruker ApexII DUO
Wavelength, Å	1.54
Resolution range, Å	50.00 – 1.60 (1.62 - 1.60)
No. of unique reflections	19443
Completeness, %	100.0 (100.0)
R_{merge}¶ or R_{sigma}¥	0.024 (0.183)
Redundancy	18.5
Average $I/\sigma(I)$	33.96
Wilson B, Å^2	19.7
Refinement statistics	
Resolution range, Å	31.47 – 1.60 (1.64 – 1.60)
No. of reflections	18397 (1404)
Completeness, %	99.9 (100.0)
R_{factor}	0.171 (0.209)
R_{free}	0.195 (0.268)
No. of non-hydrogen atoms	1297
Protein	1140
Water	131
Ion / Ligand	25
RMSD from ideal geometry	
Bonds, Å	0.017
Angles, °	1.86
Average B factors, Å^2	
Protein	13.0
Water	20.3
Ion / Ligand	10.8
Ramachandran plot	
No. in favored region (%)	104 (93.7)
No. in allowed region (%)	6 (5.4)
No. in outlier region (%)	1 (0.9)
PDB ID code	**4TRD**

‡eq. = molar equivalents (relative to 1 molar equivalent of protein)
*Values in parentheses correspond to the highest resolution shell.

¶ $R_{merge} = \frac{\sum_{hkl}\sum_{j}\left|I_{hkl,j} - \langle I_{hkl}\rangle\right|}{\sum_{hkl}\sum_{j} I_{hkl,j}}$; ¥ $R_{sigma} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}\left|I_{hkl,j} - \langle I_{hkl}\rangle\right|}{\sum_{hkl}\sum_{j} I_{hkl,j}}$

Vita

Jaime Sorenson was born in 1987 in Cleveland, OH. She grew up playing outside in the woods and ponds, igniting her interest in Biology. Over the course of her attendance at Orange City Public Schools, she furthered her education in the maths and sciences with summer programs, earning college credits at Case Western Reserve University, Northwestern University, and Harvard University. She graduated from Orange High School in 2006 with high scores in many AP courses including Biology, Computer Science, and Physics. Jaime attended the University of Rochester, graduating in 2010 with a BS in Molecular Genetics and a BA with majors in Japanese and Mathematics. The hours not spent in athletics or academics were spent tutoring fellow students and (unofficially) teaching Discrete Mathematics. She spent her junior and senior years pursuing research in mathematics based on *p*-adic numbers with Dr. C. Douglas Haessig and in biology investigating PKC isoform translocation in living cells in Dr. Coeli Lopes's laboratory at the Aab Cardiovascular Research Institute.

Jaime applied to graduate schools intent on studying developmental biology but found a way to unite her passions in Biology and Mathematics by joining the García-Moreno lab in Biophysics. In addition to research, she continued to teach as a teaching assistant and instructor for many courses in the Biology and Biophysics departments. Upon graduation in the Spring of 2016, she will spend the summer teaching high school students their way around scientific research and will share her expertise in protein engineering to undergraduates in the fall semester at Johns Hopkins University.